CREATe Working Paper 2014/4 (24th February 2014)

Archives and Copyright: Developing an Agenda for Reform

Digital Conference Proceedings

27 September, 2013 – Wellcome Trust, London

Editors:

Ronan Deazley

University of Glasgow
Ronan.Deazley@glasgow.ac.uk

Victoria Stobo

University of Glasgow
Victoria.Stobo@glasgow.ac.uk

HTTP Identifier: http://www.create.ac.uk/publications/wp000014

The Symposium "Archives and Copyright: Developing an Agenda for Reform" was held at the Wellcome Trust on 27 September 2013. It was organised by Ronan Deazley and Victoria Stobo in collaboration with the Wellcome Library, and represents the culmination of an RCUK-funded research project concerned with the manner in which the copyright regime both enables and inhibits the work of heritage institutions, and in particular archives.

A web resource offers short videos of the presentations at the Symposium, full transcripts, an introductory essay and a bibliography, as well as other project-related outputs (available at: www.create.ac.uk/archivesandcopyright).

This document is the authoritative, paginated and citable version of the Symposium proceedings (available for download at: http://www.create.ac.uk/publications/archives-and-copyright-developing-an-agenda-for-reform).

The Symposium and the production of the web resource have been generously supported by the Wellcome Trust, the UK Archives and Records Association, the British and Irish Law Education and Technology Association (BILETA), and CREATe, the RCUK-funded Centre for Copyright and New Business Models in the Creative Economy.

About this Resource

This resource was produced by CREATe, the RCUK Centre for Copyright and New Business Models in the Creative Economy (www.create.ac.uk) in collaboration with the Wellcome Library (www.wellcomelibrary.org)

Digital production team

Videography: Gabriele Cannizzaro
Photography: Francesco Dionisio
Line Production: Bartolomeo Meletti
Digital Production: Andrew McHugh and Sukhpreet Singh
Executive Production: Ronan Deazley and Victoria Stobo

The full digital resource is available from:

http://www.create.ac.uk/archivesandcopyright

Citation Guideline

Please cite this resource as:

Deazley, R. and Stobo, V. (eds) (2014)

Archives and Copyright: Developing an agenda for reform

Digital proceedings of RCUK symposium

www.create.ac.uk/archivesandcopyright

CONTENTS

PART I - INTRODUCTION

INTRODUCTORY ESSAY

Ronan Deazley and Victoria Stobo

The world in which archive services operate has changed radically in recent years. The predominance of digital records, the growing technical complexity of digital record-keeping, and the fragility of digital information all present archives with robust challenges in preserving authentic information and records in perpetuity (*Archives for the 21st Century*, 2009, pp.1, 18). The digital environment has also had a major impact on society's expectations about access to and use of information in general and archival holdings in particular. People now expect resources to be accessible online as and when it suits them, and the trend towards accessing archival records online, versus onsite access, is impossible to ignore. In 2011-12, for example, for every document ordered onsite in the reading rooms of The National Archives (TNA), 211 documents were accessed online by visitors to TNA's DocumentsOnline service (TNA, 2012).

Archivists are all too aware of the importance of enabling digital access to their collections. As Lesley Richmond, University Archivist at the University Glasgow, puts it: "Users are demanding, they're unforgiving, and more and more they are very unimpressed if archivists cannot produce or provide material online" (*infra*, p.28). Mandated with preserving records created by individuals, public and private sector organisations, and government, archivists want to make those records as accessible as possible to as many people as possible. Preservation without access is redundant. Indeed, preservation *is* access, and access within a 21st century context means digital access. The government understands this, and the archive sector has been urged to rise to the challenge of increasing the proportion of the nation's archival records that are available online (*Archives for the 21st Century*).

However, making archival material available online inevitably triggers concerns, on the part of conscientious archivists, about copyright law. Once described by Terry Cook as "[t]he perennial hornet's nest for archivists" (1984, p.21), copyright profoundly shapes the work of archives as it relates to both preservation and access. To be sure, in many respects copyright enables and facilitates the work of archivists; but it also inhibits and frustrates that work, and not just because archivists find the law to be complicated, confusing and intimidating

(Dryden, 2008, pp.230-36). Rather, originally devised for an analogue world, the law as it affects the work of archives is no longer fit for purpose. In 2009, Clifford Lynch, Director of the Coalition for Networked Information, suggested that copyright represented "the single largest and most intractable problem as far as dealing with preservation and stewardship of the cultural record going forward", and "the single biggest barrier to making available the existing cultural record, and cultural and scholarly heritage" (quoted in Korn, 2009, p.21). We would go further. Put simply, we believe the UK copyright regime is failing archives as memory institutions: it is failing and frustrating the efforts of archivists who dedicate their professional lives to preserving and making the country's archival collections as accessible and as useful as possible.

The government does recognise that there is a problem, and it has taken action. It has been consulting on reform of the existing copyright exceptions as they relate to libraries and archives since 2006, and reform is (finally) coming. New legislation is expected in April 2014, and it promises a number of positive developments. For one thing, the copyright regime for archivists should become simpler and easier to understand and apply. In addition, the scope of the current exception for preservation (*Copyright Designs and Patents Act* 1988 (CDPA), s.42) will be extended so that archives can make preservation copies of all types of copyright-protected work, including artistic works, sound recordings, films and broadcasts. (At present, the preservation exception only extends to literary, dramatic and musical works). The exceptions that currently permit librarians and archivists to make copies of certain types of work for users (CDPA, ss.37-40, 43) will also be extended to cover all types of copyright work regardless of the medium in which they are recorded. And a new exception will be introduced allowing libraries and archives to make their collections digitally available for the purposes of non-commercial research and private study on dedicated terminals based on their premises (proposed s.43A).

Moreover, both the *Gowers Review of Intellectual Property* (2006) and the *Hargreaves Review of Intellectual Property and Growth* (2011) identified the orphan works phenomenon as a significant barrier to unlocking the nation's cultural heritage, and solutions have been developed at both a national and regional level. From a European perspective, the recent *Orphan Works Directive* requires Member States to introduce a new exception to copyright enabling heritage institutions to digitise and make use of orphan work material for non-commercial

purposes. With the *Enterprise and Regulatory Reform Act* 2013 the UK government has also paved the way for the introduction of an orphan works licensing scheme that will operate in parallel to the implementation of the *Orphan Works Directive*. The UK licensing scheme is a development – the government hopes – that will allow and incentivise "archive holders to use and make available their archives" (*Impact Assessment: Orphan Works*, 2012).

These reforms and initiatives are to be welcomed. However, we remain sceptical about the extent to which they will enable archivists to realise the potential that the fullest possible online access to the country's archival holdings would contribute to local and national democracy and accountability, to learning, education and scholarly endeavour, to creative and cultural activity, and to a better understanding of identity and place for local people, communities and organisations.

Consider, for example, the developments relating to orphan works. One obvious shortcoming of the *Orphan Works Directive* concerns its scope. It does not, for example, extend to free-standing artistic works, such as photographs, maps, architectural drawings and plans. Moreover, the extent to which it will actually enable the use of *unpublished* orphan works is debatable, a profound limitation when working with archival collections (Deazley and Stobo, 39-45).

Unlike the European Directive, the UK orphan works licensing scheme will extend to all types of copyright work (including artistic works), published and unpublished, and licences will be granted for both commercial and non-commercial use. However, licences will be conditional on payment of an upfront fee. The nature of the fee will be determined by the independent licensing body to be established under the scheme, but the most recent guidance from government makes clear that '[t]he scheme *will not allow anyone to use copyright works for free*' (IPO Guidance, August 2013, emphasis added). That is, at present, it is anticipated that a licence fee of some kind will be levied for every type of licensed use, whether commercial or non-commercial. But given the nature of archival collections – largely unpublished records and documents with very little intrinsic commercial value, the very existence of which rarely turns upon the promise or expectation of future commercial exploitation – why should archives be compelled to treat these materials as commodities at all? And should increasingly stretched budgets really be spent on licence fees to make this material more publicly available?

What's more, both the European and the UK solutions to the orphan works phenomenon are contingent upon conducting a diligent search for the relevant rightsholder prior to use of any orphan work. Engaging in diligent search seems like a reasonable condition to impose when dealing with rights clearance issues on an individual basis, or for small numbers of works; but it is likely to render both schemes utterly redundant when contemplating the kind of mass digitisation initiatives that have been called for by government.

In short, while we recognise that the new copyright settlement will better enable and facilitate the work or archivists, we also believe it will continue to inhibit and frustrate that work in important ways. With that in mind, the premise underpinning the Symposium that led to the production of these proceedings was really very straightforward: what can be done to make the work of archivists easier and simpler, more efficient and more effective? Within the copyright domain, what *evidence* can be gathered and what *strategies* can be developed that might help make it easier for archivists to preserve their collections appropriately, and to make those collections as open and as accessible as possible?

When Prof Ian Hargreaves delivered his influential review of the intellectual property regime, his first recommendation was that '[g]overnment should ensure that the development of the IP system is driven as far as possible by objective evidence' (2011, p.8). Understanding what constitutes reliable and relevant evidence is not uncontroversial (in general, see Kretschmer and Towse, 2012), but there is no doubt that in recent years the concepts of cost-benefit analysis and evidence-based policy have taken centre stage in shaping copyright policy and reform. So: how can or should the archive sector respond to this development? What data can be gathered that speaks directly to the work and needs of archivists? When archives clear rights for digitisation initiatives, how often do rightsholders refuse permission? How often do they ask for payment or negotiate a licence fee? If rightsholders refuse permission, why do they do so? Does it concern copyright, or are there other non-commercial imperatives at play? If copyright fees are requested, do publicly funded archives have budgets to pay those fees?, or are they more likely to simply omit the material from the digital resource being developed? Have any UK-based archivists or archives ever been sued for copyright infringement, or even threatened with copyright litigation?

Much could be done to gather information and evidence specifically relevant to the archive sector to help policy makers and legislators

better understand the realities of the copyright regime as it affects and impacts archives, as distinct from other heritage institutions such as libraries, museums and galleries.

As for strategies, one possible strategy for making the life and work of archivists easier and simpler, more efficient and more effective, is to adopt a risk-based approach to clearing rights, rather than one of strict copyright compliance. When the Wellcome Library began to develop *Codebreakers: Makers of Modern Genetics*, it adopted just such an approach. Wellcome's vision for *Codebreakers* was to make archive material relating to the history of genetics and genomics freely available online, including material not just in the Wellcome's collection but in other archival collections also. At the time of writing over 1.6 million pages of material is now available to view through the Wellcome Library Player, all of which is made available under an attribution non-commercial creative commons licence. Recognising that the costs and demands of strict copyright compliance would seriously compromise delivery of the *Codebreakers* project, the Wellcome Library adopted a risk-based approach to clearing rights, and without doubt, this strategy contributed significantly to the success of *Codebreakers*.

In our opinion, a risk-based approach to copyright compliance is a strategy that merits serious consideration and review on the part of the archive community, and these proceedings represent an opportunity for doing just that: for reflecting upon the merits and problems of adopting a risk-based approach to copyright compliance within an archival context.

INTRODUCING THE SYMPOSIUM

Archives and Copyright: Developing an Agenda for Reform was organised around four related panels, followed by an open discussion with the audience.

The first panel introduced the Wellcome Library Digitisation Programme and its pilot project, *Codebreakers: Makers of Modern Genetics*. Speakers included: Christy Henshaw (Digitisation Programme Manager at the Wellcome Library); Caroline Herbert (former Content and Metadata Officer at the Wellcome Library) who was responsible for tracing and contacting the rightsholders identified as medium and high risk during the *Codebreakers* project; and Lesley Richmond (Archivist and Deputy Director of the Library, University of Glasgow). The

discussion explored why and how the Wellcome Library developed their risk management strategy for making copyright-protected material available online, as well as the practicalities of clearing rights adopting a risk-based model, and of collaborating in a mass digitisation project as a third-party partner archive.

A number of key insights emerged, including: the extent to which concerns about data protection often overshadow copyright compliance issues when making archival records available online (a burden that rarely arises when digitising library collections); the very high percentage of permissions granted by rightsholders contacted by the Wellcome (as well as their overwhelmingly positive reaction to being included in the project); and the importance of discussing and negotiating the future use of a collection – from a copyright perspective – as part of the deposit agreement, and *before* an archive takes custody of the collection. Perhaps most significant is that the Wellcome Library have made considerable progress in developing a sustainable risk-management strategy to copyright compliance, one designed to remain relevant in the long-term but flexible enough to evolve in step with legislative change, as well as one that will allow the Library to progress its digitisation agenda in a pragmatic and sensible manner.

Victoria Stobo was the principal speaker on the second panel, presenting results from her six-month study of the *Codebreakers* initiative. The study (*Copyright & Risk: Scoping the Wellcome Digital Library*, 2013) examines the merits of, and problems encountered by, taking a risk-managed approach to rights clearance, and draws upon data collected at the Wellcome Library, as well as interviews with key *Codebreakers* project staff at the Wellcome and the partner archives.

Stobo's presentation addressed three areas that feature prominently in the study. First, the results of the rights clearance process (that is, the number of permissions received, refusals received, non-responses and orphan works identified) were considered in relation to other cognate projects and studies. Second, the policies and processes developed by the Wellcome to manage the digitisation and clearance process, including their risk criteria and takedown policies, were explained and discussed. Third, in analysing the outcomes and the lessons of the *Codebreakers* initiative two issues were singled out for specific comment, relating to *communication* and *reputation* respectively.

On communication, the manner in which the Wellcome Library engaged with rightsholders (explaining both the non-commercial nature of the *Codebreakers* project as well as their robust policies for managing

sensitive data and dealing with takedown requests) was considered integral to the success of the project. On reputation, during the interviews with staff it became clear that the main risk factor for those involved in *Codebreakers* concerned reputational damage and not the financial risks associated with potential copyright litigation (a scenario regarded by interviewees as highly unlikely). In many respects, it is not surprising that reputational damage was the preeminent concern for the institutions involved; for all archives, maintaining a reputation as a trusted, reliable repository is imperative in managing relationships with depositors old and new. And while this concern over reputational damage was linked to making copyright-protected material available online without express permission, it was even more strongly associated with the potential consequences of disclosing sensitive personal data in contravention of data protection legislation.

For the third panel, the Symposium's focus shifted to the current proposals for copyright reform within the UK as they affect the archive sector. The Intellectual Property Office was represented by two speakers: Robin Stout, a Copyright Policy Advisor, and Nick Munn, the Deputy Director for Copyright. Together, they presented an overview of the recent work of the Intellectual Property Office in realising the recommendations of the *Hargreaves Review*, while specifically addressing the proposed changes to the *Copyright Designs and Patents Act* 1988. In particular, Munn emphasised the IPO's willingness and desire to engage with informed stakeholders and constituencies – such as the archive sector – about the nature and implementation of copyright reform, and stressed the value and importance of providing the IPO with reliable data about the reality of copyright clearance and diligent search within an archival context.

Tim Padfield, the former Copyright and Information Policy Officer at TNA, concluded this panel with a critique of the proposed reforms. A number of the reforms were welcomed, such as the expanded scope of the preservation exception, and the steps taken to simplify the operation of the legislation. In addition, Padfield drew attention to the potential significance of s.76 of the *Enterprise and Regulatory Reform Act* 2013, a provision that enables the government to reduce the duration of copyright for very old unpublished works which, because of the current transitional provisions on the copyright term, remain in copyright until 2039 (CDPA, Schedule 1, s.12). Government action on this issue would not only lead to a significant reduction in the number of orphan works held by UK heritage institutions, but it would, in

Padfield's words, resolve one of the major absurdities of the current copyright regime.

However, Padfield also called for further action in this domain. For one thing, he reiterated the need to extend the scope of the current exceptions that enable copying of published work by libraries (CDPA, ss. 38 and 39) so they also apply to archivists. For another, he remained sceptical – as we do – about the usefulness of the orphan works schemes for large scale digitisation initiatives, leading him to the conclusion that risk management would become an increasingly important issue for the archive sector as a mechanism for dealing with rights clearance. And in some respects, Padfield's observation on this point was mirrored by Nick Munn. While making explicit that the government could not condone a risk-managed approach to copyright compliance, there was, nevertheless, a frank acknowledgement and appreciation of why archivists might feel compelled to employ such a strategy as a practical response to the challenges of enabling greater online access to archival material. Munn put it as follows: "We know why you take it [a risk-managed approach], we know why you're interested in doing that for your own laudable objectives, and that will continue to be an option, as far as I can see, into the future …" (*infra*, p.69).

In the fourth and final panel the opportunity and scope for mass digitising archive material was considered from the perspective of US copyright law. Professor Peter Jaszi (American University) and Professor Matt Sag (Loyola University) spoke to the long-running litigation over Google Books, and to the decision of *Authors Guild* v. *HathiTrust*, a case that developed out of the Google dispute. In *HathiTrust*, in late 2012, District Judge Harold Baer Jr held that the mass digitisation of library books to enable access by persons who are blind or visually impaired, and to facilitate certain non-expressive uses of those materials (for example, for data or text mining purposes) did not amount to copyright infringement; rather, those activities constituted fair use under s.107 of the US Copyright Act 1976. Moreover, shortly after the Symposium, on 14 November 2013, Circuit Judge Denny Chin dismissed the lawsuit against Google, accepting their argument that scanning more than 20 million books and making 'snippets' of text available for online searching fell within the doctrine of fair use.

While both cases specifically relate to the digitisation of library material (published books), and both are likely to proceed to appeal, nevertheless, the potential implications of these decisions for the

archive sector within the US are obvious. Because the fair use doctrine is open-ended, flexible, and situational, the justifications endorsed by the courts in these decisions are, as Jaszi puts it, "merely examples of the kinds of justifications that could be advanced successfully … for a variety of other mass digitisation projects" (*infra*, p.86). Again, in Jaszi's words: "[f]air use can and does reach activities like preservation and access, which are at the heart of the archival enterprise and in particular, at the heart of the move toward both digital preservation and the provision of remote access to archival collections, and it can do that without regard to the specific kind and character of the copyrighted material involved, or even the exact nature of the institution engaged in archival activities" (*infra*, p.84).

The fourth panel concluded with Peter Hirtle, Senior Policy Advisor to the Cornell University Library, Research Fellow at the Berkman Center for Internet & Society at Harvard University, and the author of *Copyright and Cultural Institutions* (2009). In turn Hirtle considered and dismissed adhering to a policy of strict copyright compliance or hoping for archive-appropriate law reform as strategies that were unlikely to provide a meaningful solution to the challenge of digitising archive collections. Instead, drawing upon themes from Jaszi and Padfield's contributions, he advocated greater reliance upon fair use in justifying archival digitisation initiatives (for US-based archivists), as well as a more general recalibration of the global archive community's attitude to risk and the culture of copyright permissions. In short, Hirtle's take-home message was that, in managing the challenges of the transition to the digital, archives should let their mission "rise to the fore" (*infra*, p.99). To be sure, copyright should be respected, but so too should archives respect their mission to enable access, to educate, to promote the use of historical materials, and to encourage new scholarship.

* * *

Before turning to the transcripts of the Symposium, we would like to take this opportunity to acknowledge that without the involvement and cooperation of the Wellcome Library it would not have been possible to realise this project. We thank them for inviting CREATe to reflect upon *Codebreakers* (from a copyright perspective), and for offering us a platform for locating those reflections within a broader national and international debate about copyright policy and reform.

We regard the Symposium, and these proceedings, as part of a necessary and emergent conversation about copyright and archives, and we welcome further comments and contributions in advancing that debate.

PART II – EDITED TRANSCRIPT OF PANEL DISCUSSIONS

WELCOME

Richard Aspin

Head of Research and Scholarship at the Wellcome Library

RA: I'm Richard Aspin, Head of Research and Scholarship in the library here, AKA Head of Special Collections, which is a job title which is perhaps a bit more familiar in the context of most institutions. It's my job really to do three things, firstly and perhaps most importantly, is just to welcome you here and say how nice it is to be able to host this event, and I hope you'll have a very interesting and stimulating day. The second thing is to make sure you're comfortable and safe throughout the day, so I just need to remind you of where the emergency exits are and the loos. Sue Davies, my colleague is showing you where the fire evacuation emergency exit is at the moment, and the loos you may have already encountered, which are down there, cunningly hidden behind the wooden walls in this rather deceptively mysterious building.

That's all fine. I wanted to say a few words just to set the scene for why this event is happening here, and why Ronan and CREATe rather flatteringly think that our Codebreakers project is of interest. The Wellcome Library is quite a strange thing - it's a library which is often described as the library for the history of medicine, but it's really much broader than that. We have hugely varied collections, and we have amongst those collections - which range from the ancient world through to modern times - we have very important biomedical and other archives. And rather perversely, we decided when we wanted to begin mass digitalisation of our collections to make them more accessible online, that we would begin with perhaps the most difficult body of material, the modern and contemporary archives with lots of letters written by people, most of whom are likely still to be alive, some even possibly quite young. So, why did we start doing that? Well, we

have been tasked by our governing body to collect and develop our collections around [five main themes](#), the first of which as you can see is genetics and genomics, so we began to create an online resource based around archives of genetics and genomics. We are going to move on, we hope, eventually to other areas of activity. This is the [homepage of the Codebreakers site](#) which we developed as a pilot for mass digitisation of contemporary materials. One of the main challenges we had with this material, which is obvious from what I've already said, is the issue of clearing copyright from many of these copyright holders represented in the archives. And so I think without further ado, I shall hand over to Christy Henshaw, our Digitalisation Programme Manager, to take the story on from here.

PANEL 1

INTRODUCING THE WELLCOME DIGITAL LIBRARY: CODEBREAKERS AND BEYOND

Speakers

Christy Henshaw (Digitisation Programme Manager at the Wellcome Library) – hereinafter **(CHen)**

Caroline Herbert (former Content and Metadata Officer at the Wellcome Library) – hereinafter **(CHer)**

Lesley Richmond (University Archivist and Deputy Director, University of Glasgow Library) – hereinafter **(LR)**

Chair: **Ronan Deazley** (Professor of Copyright Law, University of Glasgow) – hereinafter **(RD)**

Questions and Answers

Chris Bird (Senior Legal Counsel at the Wellcome Trust) – hereinafter **(CB)**

Susan Corrigall (Copyright Officer at the National Records of Scotland) – hereinafter **(SC)**

(CHen) I'm going to tell you a little bit about how we approached the problem of copyright in archives. We didn't just have a problem with copyright in archives, we also had a problem with copyright in books, and we had a problem in some other things that we were doing. But this conference is about archives so that's what I'm going to focus on today, but some of what I'm going to talk about applied across the board to all of our different collections that we're digitising.

I manage the digitisation programme, I've been here for about six years, and over the past three years was when we decided to really scale up digitisation. Copyright suddenly became not just something we had to worry about, but also something we had to seriously resource if we were really going to do it.

The vision for Codebreakers, as Richard has already said, was to make history of genetic material available online from our collections, as well as bringing in content from outside sources. So the archive was mainly comprised of late nineteenth, not very much of that, mostly twentieth century published and unpublished material. It's quite a large collection, so we've already put about a million pages into our repository and we're going to carry on adding to that, as we carry on digitising. We want to make it all freely available under an attribution non-commercial licence, if possible, and we want to ensure long-term access and preservation - those are the key goals of our digitisation project.

The archives come from us as well as five other contributors, that's [UCL](#), [Glasgow University](#), [Coldspring Harbour Laboratory](#) in the States, [King's College London](#) and [Churchill Archives Centre](#) at Cambridge University. The kinds of things you get in an archive, which I'm sure you will all know or most of you will know, is a lot of different kinds of things and you don't always know what you're going to get until you open up the box. So obviously correspondence was one of the big areas for copyright clearance. It was one of our big concerns about how we would deal with that, because you have letters coming in, letters going out and what is the copyright on them, it can be quite confusing, as well as all this other stuff.

So what did we do? Well, first we had to start talking about it, thinking about it and consulting with ourselves and with others, and some of the things we had to think about were, what are the real risks here? What is the law? What isn't the law? How does the law relate to the material we're digitising? What is our appetite for risk? How much risk are we willing to take? What kind of contingency measures might we want to put in place because we know we're going to take risks? What's the criteria going to be to instigate copyright clearance? We knew we would have a take-down policy - what would that look like? We went through internally, but we also have a programme board for our entire digitisation programme with external members on it, and they review and approve a lot of the things that we do.

As we went through that process, one of the things we had to do was assess the risk; what is the real risk? I've put, copyright clearance is about managing risk, always. I think it's really rare you're going to have a situation where you're one hundred percent sure that there's no risk. I'd be very surprised if there's never any doubt.

What are the risks? We decided there were two main risks for us. One was financial, and that could be the cost of a lawsuit, or the cost of removing material from the website - you might as well factor that in, that takes time. We thought that was a pretty low risk, because of the type of content we were digitising, in terms of the archives. The content in the archives is not produced for any kind of financial gain for the most part, it's older, it's still in copyright, but it is older. There were a lot of reasons why we thought that, financially, we didn't have a huge risk there. We thought reputation was a higher risk, that we'd be seen as overstepping our bounds, we'd be seen as breaking the law, we'd be seen as taking advantage of people, and that could result in a loss of trust by depositors who want to leave their archives with us, our peers and other key stakeholders. Again, mitigating risk covers both those areas, but the reputational risk was the one that we thought about the longest and the hardest.

We had to take action at some point. We had to actually do something. We couldn't just keep on talking about it all the time, we wanted to digitise things. We knew we couldn't do nothing, but we knew we couldn't to everything, we couldn't trace everyone. We knew we needed to collaborate, because we don't know everything. Archives, for example, coming in from outside collections, the people who own those or hold those archives are going to know more about what's there than we are, and we needed to get their input on that. We had to make judgement calls, so we couldn't really define a list of rules that said, absolutely this must be cleared, and absolutely this doesn't need it. We had to make a judgement call, and we needed simple criteria to do that, otherwise we'd just tie ourselves up in knots and never get anywhere.

We also had to think about proportionality, how much effort are we putting into it compared to the risk, compared to all these other things

that I've written here. And I'm going to explain these just a little bit because I think they're really important.

One is the level of risk, which is subjective, depending on your appetite for risk. And then you have other things, like the extent of the metadata. How much detail have people gone into when they've been cataloguing this stuff? Have you got a box of letters which is just catalogued as a box of letters, and you have no idea what's in it? Or have they catalogued every single letter and said who they're from, who they're to, who they've mentioned and what they've talked about? You could go from one extreme to the other, most are kind of somewhere in the middle. Our metadata tends to be towards the lighter end, so you can't go through and just scrape this for every single name, even if you wanted to, which we wouldn't necessarily. Even if you have lots of metadata, how is it structured? Did someone just write a list of names, so that you find it in a search term, or do you know that this person wrote this letter and that person received the letter, or that person was simply mentioned in the letter. There is that element to the data as well, which can make it more or less easy, and this applies to sensitivity as well, which is an area we also had to look in to - judging what you've got in your collections and therefore, what you can do with it.

Then multiply all of that by the number of copyright holders you think you might have, or you know you have, and you end up with the potential effort that's required to copyright clear the archive. We probably had thousands of copyright holders potentially in these archives. There was no way we were going to be able - or we felt we needed, to be honest - to identify every single one, and to trace every single one, which takes quite a lot of research and quite a lot of time, as you will hear from Caroline after me. We created shortlists for each collection and we cross-checked between them, because a lot of our collections are related, so you have content from the same people in multiple collections. Then, once we had a shortlist of names that we wanted to trace, that met our criteria, which was that they were higher risk and we should try to trace them, because either it was for profit, creative product, there was a literary estate for example or something like that. We had a series of criteria and I think your criteria is one of

those things again, I can't give you a formula for, it's going to be related to your appetite for risk. We had one that we used. We found that of the names that we chose to trace, we were able to find some kind of address for, or email for, eighty four percent of everybody. Of those, seventy eight point five replied, so most of them did actually reply, and of those who replied, only two said no, and I don't know if they were about copyright, because I don't think they said. They just said, "No Thanks." And that's quite common, we found with the books for example, and other copyright clearance, often people just say no thanks, they don't really say why.

We felt that this strategy was sustainable, because we were able to reduce the effort down to something that met our criteria, it was something we felt we could resource. It did take time and it did take effort, and we did have to spend money doing it, but we felt like this was something we could carry on doing in the long-term, and I think that's actually quite a big success, because we could very easily have gotten to the end of this and said, "Oh, we can never digitise any more archives, it's just too hard," and we haven't done that. And the proof is in the pudding, we've only put it online very recently, so what happens, you know. We've done all this, we feel like we've achieved something, we feel like we've done a pretty good job, but what happens if somebody does come to us and says, "Actually, that's my copyright, I don't want that up on the website." The first thing we do, for whatever reason, it could be a sensitivity reason or a copyright reason, or they might not even give us a reason, we would take it down. Then we'll think about it, we'll review it, we'll get back to the person if we think we need to, but if it seems like it's a legitimate request and they really are the copyright holder and they have the right to ask us, then we'll just keep it off the internet. We don't want to upset people, and that is the URL for our take down policy if you want to look at it. I thought instead of going through it, because I don't have a huge amount of time, you could just look at it in your own time. If it's a copyright reason why somebody wants something taken down, we will take it down, but it doesn't mean we're deleting it necessarily from our systems, although we could do that if required. The copyright will expire, and a lot of the

stuff we have, the copyright will be expiring very soon, and then we can put it back online.

So, that is the end of my presentation, and I'm going to move on now to Caroline, who did a lot of the copyright clearance work.

[Applause]

(CHer) As Christy said, I was doing a lot of the donkey work on this, so I'm here really to fill out the outlines of how we put all those policy decisions into practice. I guess it's easy to make it sound like it was just like going from A to B. Once you made those decisions, you just get on with it, but I think it was a bit more complicated than that.

First we had to decide who we were going to ask for permission, and how we were going to narrow down those thousands and thousands of names in the metadata, into the list of people we were going to trace. The criteria focused on how well known someone was, if they have a high public profile, or if they had died and there was a literary estate. But all those things, they require a certain level of subjectivity. So how do you decide if someone's well known, or had a high public profile just from looking at the name? You need people who actually know a bit about the subject, or know a bit about the history of genetics as it was for us. So, even in the Wellcome Library, there were several staff involved at different stages in the discussions of pulling together this list of names. Again, Christy referred to the problem of the metadata - that if there's a list of names in the metadata it can be that letters that were addressed to people, or photographs of them, or material about them. You can't just go through and do a computer search, pull out a list of names and say "Ah, these are people we need to put on our list to trace," because there's no guarantee there's anything that's in their copyright. You don't want to end up spending a lot of effort tracing someone, only to find out that there was nothing they created in the archives.

That involved looking at all the records at the individual level. So, if you had someone's name in a record you had to look at the context, you had to read the file description, and it doesn't sound like it takes long, but

when you multiply that for all the thousands and thousands of records for the material, it does all start adding up.

As we were doing this, we were also liaising with the other institutions who were contributing to the project, and really the Wellcome Trust was at the hub of this activity. We kicked it off and an initial list was circulated to the other institutions, and then they checked the names against their collections and also added in any names that they had found that they thought ought to be included based on the criteria. That was fed back to the library and then the additional names were checked by us, re-circulated to the other institutions, and there was this iterative process going on which again took several months to go from the start of the discussions, down to this final list of names - these are the people we're going to try and trace.

That was the first stage, now we've got to find some addresses for these people. Yes, it can be very easy if someone's still alive, they have a high public profile, they're in Who's Who, you can get several of them knocked off before your first cup of tea in the morning. If they're dead and they have a literary estate, again, that's fairly straight forward and there are databases you can use and so on. But if someone's died, and there's no literary estate, then you have to try and work out who the heirs were, who's the next of kin, who you think the copyright holder will be, and that's where the detective work started to come in; going through, trying to find out if someone had children, or a spouse, or siblings. Again, you can get some of this from sources like [Who's Who](#), or obituaries will tell you - but if someone died several decades ago, even once you've got their next of kin, you're then going through the whole process again to see if they're still alive, and if they're not, who is the rights holder? Even when you've settled on the name, you think that this is the person who currently holds the rights for this material, you then have to try and find the address. Just because someone's father or grandfather had a high public profile and was in Who's Who and was very famous, it doesn't necessarily mean they are. Sometimes in all of this, we found that we'd just been heading down dead ends, and we got to a point where we'd have to treat these rights as orphans. We think we've done all we can do, there's this whole idea of a due diligence

search, but without any real guidelines - in the library we had sat down, we'd said these are the things we think we need to do, these are sources that we'll check for everything, and we kept records as we went. We've tried looking there and that was no, and we tried looking there and they're not in there, and we tried this and they're not here and we looked up this, and we did this. So we have all those records, but there's nothing that we can measure those against, and know we've done enough when we got to this point, and if we do more then it's just wasting time and money, because we're looking for something that isn't there. There's just that uncertainty there.

But the flip side, as Christy said, we did manage to trace most of the rights holders we were looking for. Then we started sending out letters and asking for permission to make the material available online, and the replies starting coming back in, which was great. There were lots of people with positive responses, people saying "Yes, very happy," signed permission forms rolling in. I went off on holiday and came back to a pile on my desk, took me a while to go through them.

There were a few rights holders who went beyond that and they wrote to us, they got in touch, they emailed or phoned and said, "Brilliant, I'm really pleased you're doing this," "I think it's brilliant, I'm really looking forward to this material being online, it's great that you're making it available," and, "When is it available? Can you let me know when the site launch is, please? I want to have a look, I want to go through." There was that sense of real positivity coming back from these rights holders, that it wasn't a matter of indifference to some of them, that they were really keen to support the project, they thought it was worthwhile.

Now, not everyone was that positive, I'm sure you're thinking, surely not everyone was that enthusiastic? Several people did get in touch when they received our request. The initial request letter generally went out with a sentence or two to try to describe the material, but people would say, "Can I see copies of these letters? What's in them? Is there sensitive material? Is it going to be embarrassing? Is it going to release skeletons from my family closet?" Obviously, that was absolutely no problem, we were very happy to make these copies, although there was the issue of the time-consuming nature of going

through several files and finding that individual letter in each file that the person had written to Francis Crick, or whoever. I was always very grateful for the people who had filed their correspondence in alphabetical order. It took time but it didn't take as long as it might have done.

But people who asked for copies, it was clear that their concern wasn't actually about copyright, because if it had been, they would just have said no when they go the letter; not interested, don't want to get involved. It was the sensitivity, that was what was bothering them - what was in these letters? We'd send the copies off and they could see that they were scientific but professional, we had the sensitivity policy to make sure that nothing went online that was going to be sensitive to living descendents, or was going to breach data protection or anything like that. The rights holders would get the copies of the letters and say, "Yes, that's fine." No problem, very happy.

And in all of this, money hardly ever even came up in discussion. Generally, the rights holders would send the permission forms back and that was the end of it. Obviously, it's a non-commercial project, the material's being made available for free, so that may have had an effect, but we weren't spending hours trying to negotiate on money, it just wasn't an issue for anyone.

Now we're at the end and looking back, what are the key points that I think we've learned? It's clear that even when you've got the simple, straightforward process, it's time consuming to go through, but at the end of the day, the vast majority of rights holders are very happy for the material to be displayed online. As I said, all the concerns that were expressed to us, and even in the refusals where there was a reason given, again, it was all about sensitivity and not about the copyright. I think those are points which can feed into future projects at the Wellcome and elsewhere, and maybe go some way to answering some of the questions Ronan was asking at the beginning. Thank you.

[Applause]

(LR) Morning, I'm Lesley Richmond. I'm still the archivist at the University of Glasgow, but I'm also now the Deputy Director of the Library where copyright is another issue altogether. But this morning I'm here to reflect on the experience of the third party archives involved in the Wellcome Digital Library project, and there are quite a number of them in the audience. Would you like to make yourselves known if you're part of this project? But apologies to them, because really what I'm going to talk about this morning is more about Glasgow's experience. But we all took part in a grand project to try to prove proof of concept of the Wellcome's new venture creating a digital library of material held in its own, and other organisations' archives.

At Glasgow we were, on the whole, the odd ones out because we held little material on the Codebreakers, so the Cricks, the Watsons and the Franklins. Two of our collections were from people who were medical geneticists, and we knew from the outset that our main challenge was going to be around the digitisation of personal sensitive data, or the non-digitisation of personal sensitive data. [Malcolm Ferguson-Smith](#) was one of the first geneticists to provide a diagnostic and counselling service to patients, so you can get an idea of what I might be talking about.

But we were very keen to be involved in such a project, because we realised that users, and especially young users these days, expect archive collections to be made available online. You only have to witness a student constantly pushing buttons of their computer because they want the description of the item that they want to see to resolve into the document itself, and they just think they system's not working when that doesn't happen. Users are demanding, they're unforgiving, and more and more they are very unimpressed if archivists cannot produce or provide material online.

I was quite incredulous when Wellcome said it was going to be medical genetics as the project. I did, I think, actually laugh out loud over the phone because I could see some major stumbling blocks over research value, value for money, copyright and data protection. It soon became clear why medical genetics had been chosen. The Wellcome had [Research Resources in Medical History](#) themes, medical genetics had

been the last theme, and we needed metadata, a good catalogue, so it did make sense.

It became clear quite soon that, although we were well used to working with Wellcome under the Research Resources and Medical History scheme, we were now dealing with a new project team and that had its own challenges in its own way, because we had to build new relationships with a new team. But it's exactly the same sort of relationships you have to develop when you're working with any multi-interdisciplinary team when you're doing a major digitisation project. It's mainly won over understanding terminology. But once we'd got our items and our pieces worked out, we worked well together.

Our main concern was the subject matter, the medical genetics. Genes go on well beyond the normal hundred years' lifetime of a person. Genetic disorders run through families, we presume forever, I don't think we actually know yet, so we could see problems and we were entering uncharted ethical areas.

Our first concern was more about data protection than it was about copyright, and that remained our major concern throughout the project. We were also concerned about the robustness of the metadata, i.e. the cataloguing, because we had not catalogued the two collections; they had been done by the [National Cataloguing Unit in Bath](). We just didn't have the same understanding and control of the collection that we would have had, had we catalogued it ourselves. We were unsure what appraisal had been carried out, we were unsure about the descriptive cataloguing, we were unsure about the data protection assessment that had taken place, and we couldn't always understand the arrangement.

We employed a full-time project manager who ironed out some of the cataloguing issues, and monitored the work rate of the digitisers, but whose main job was to undertake sensitivity checks of all the documents, and also undertook the copyright assessment.

At Glasgow we have data protection assessed at the item level, deliverable to the search room level, but that's as much a warning about

the material. Some of the material in the file or the series may not be able to be produced to any readers in the search room, but the chances are that, if it would be useful for that person to see just parts of the file, it could be given to them. And certainly in the case of our medical geneticists, we did know that there were scientific documents and articles in the files which would be of great benefit to researchers, and these would not breach data protection, which is why we did do sensitivity down to document level.

Although initially Wellcome told us it was a project where everything was going to be digitised, that is certainly not what we did at Glasgow. We did not digitise anything that contained personal sensitive material. We could have done that and had it closed for twenty to thirty years, but we just decided we were not going to do that. We also did not digitise any published material that was illegible; that was a value for money judgement. And we did not digitise any published material that was not annotated.   We did digitise the front page so that any researchers would know what article the person concerned had thought was valuable at the time, to put into their research file.

The reason for not digitising anything that contained personal sensitive material was threefold. First of all it was value for money; who knows how we shall digitise anything in twenty or thirty years' time? I hope it will be easier and quicker by then. The second one was reputational risk. If the material was digitised and handed over to a third party, would it remain closed for twenty to thirty years? Perhaps we were being overly cautious, but we just did not want the material to go out of our control, be held somewhere else, where somebody at some point could just flick a switch and suddenly material would be all over the internet, and that's no disrespect to Wellcome at all, we just felt safer if it was with us. The third issue was ethical consideration; would after being closed for twenty to thirty years, would we be able to open it ethically, and make it available for the world to see? Just what are the personal sensitive issues involved in genes? Genes live on in families and we had material that could reveal information about future generations with distinctive names and we decided we would wait and see what gene development takes place in medicine. I'm sure in the

future we'll all know what our genes do to us, exactly when we're going to get our next dose of the flu or whatever, and we might have a different more relaxed view to all this but at the moment, we just don't know.

But I do need to say something about copyright. All the third parties involved in the project had their copyright understanding dramatically increased by taking part in the project and by talking and discussing the issues that came up. The risks for the third parties became less of a concern when we understood that it was Wellcome that was publishing the material. They would be the ones that would be liable and we had that understanding in the contract, hence why, as a contributor, [Glasgow University Archive Services](#) was very happy to be producing the digitised material under contract, and we've been very, very strident in our recommendations that future projects should also always be under contract rather than a grant. Contracts are difficult to get your institutions to engage with, but they're worth the effort.

It also became clear that copyright risk assessment varies from institution to institution, because every institution has got different relationships with their copyright holders; a different relationship with the material that you hold, and there is a context of relationships and ownership. For us, our higher risk end was over the copyright holders we were aware of, with whom we had a relationship, and we understood that Wellcome would not have that relationship at all with those copyright holders, and they obviously had different relationships with their copyright holders.

One of the main things that all the third party archives were able to do early on was to obtain copyright clearance from the depositors or the creators of the collections that we held, and on the whole that could release, if you get clearance, about fifty percent of the collection in the first instance, at least for the material that the person concerned created. Again, that's one of the recommendations that should go forward, that all archivists, when they obtain or are gifted a collection, should insist, with all those long lists of things to do when you're talking with the donor, is to have copyright, an understanding about copyright written into the contract. It's as difficult as it's been

described by Caroline for archivists to even find out where the descendants of the owner of a collection now resides anywhere in the world.

It was excellent that the whole administrative burden of trying to find out who the copyright holder now was, was taken on by Wellcome, so it really was a joint collaborative project. There is the question to be asked, about the list that Wellcome now has with the contact details of the rights owners. There isn't enough resource to support that to maintain it, as far as I'm aware, going forward. The project clearance was for individual research purposes but users in the future may well want to ask to use the material for commercial purposes, for publishing in books, and they will also need to seek permission to do that from the copyright holder. It's a huge task, it's perhaps impossible to try and keep this list up to date, but now we've got it perhaps we should be trying to do something with it. We've invested a lot of money in it and perhaps we should be keeping that asset live.

Codebreakers was an important project for archivists to work out protocols for digitisation of twentieth century collections, collections that are rich both in copyright and personal sensitive information issues, and it has produced good working practices and guidelines for other projects to follow, and you'll find a lot of that on the website. But my hope for today is that we can put together an agenda to influence reforms of copyright law, so that archivists can, and are able, to fulfil their main mission of making the archives and records of everyday folk going about their everyday business, people and institutions, accessible.

Twenty first century accessibility means being available online, it needs to be made digital and that does mean that you have to copy material, so what I'd really like to see out of today is something going forward that ensures that copyright makes the job of the archivist possible, and no longer impossible. Thank you.

[Applause]

(RD) So, you now know who told me that copyright made their job impossible. One of the things I wanted to do when I was organising this event with Sue Davis and Wellcome was to take huge advantage of Wellcome's excellent catering facilities, so we've folded in as many tea breaks and coffee breaks as I thought was decent. We have a couple of minutes before we're going to break for tea and coffee and then we come back at 11:15. There will be a Q and A session at the end of the next panel after Victoria Stobo speaks, and all of our current panellists will also be involved in that Q and A session, so there will be opportunities both to talk to them over tea and coffee, and ask questions at that stage, but perhaps if anyone has a burning question that they must ask now, before we break for tea and coffee, I'm quite happy to take some of that. We do have some roving mics and if you've got questions throughout the day, please self- identify yourself and the institution that you represent before you ask your question, and that will help in terms of producing a transcript and producing a set of digital proceedings.

(CB) Chris Bird, Senior Legal Counsel at the Wellcome. I've got a question for you and I'm sorry I don't remember your name and I don't have a programme because I came in late.

(LR) Lesley.

(CB) So I got the very strong impression that you were initially sceptical, subsequently relieved about putting the contract in place about the fact that the copyright issues were dealt with to your institution's satisfaction, but I still get the sense a little bit that you feel that the copyright issues remain and the fact that your institution's copyright issues were dealt with you're happy with.   And I just wondered what your sense was at the end of this was whether we've just been lucky that we haven't had more in the way of take-down and objection or whether this was a well managed process and is reproducible with other projects.

(LR) It was a well run project and the processes that have been put in place can have the same results elsewhere. I think from what Caroline said as well is that most people don't understand copyright anyway and

they're astonished that you have to ask them permission to do this. Those people who understand copyright for commercial purposes and that's where I now know I've got a very large appetite for risk [laughter] but if you have spent a lot of money creating something then you should be protected from that, I would agree with that. But archives on the whole are material that were not created themselves for commercial gain, it's just you going about your normal business. So copyright… it was comforting that Wellcome has taken the liability so that's a relief and I think this project is a way to demonstrate going forward that no damage has been done so can we please change copyright law.

(CB) Okay thank you that's useful and we were very concerned about copyright liability and we didn't know what our risk appetite was either. We just formed a view as we went along.

(LR) That's it you just know that copyright exists and then you speak to people who are professors of copyright law and you get more concerned. It's very good when they suddenly say we must do something about this for you. So more copyright lawyers should become archivists, as a career break, then they'll understand.

(RD) As a Venn diagram there aren't many people sitting around [inaudible 51:14], not in the UK. Thanks very much for your question. Anyone else? Yes, we've got one at the back and then we'll break for tea and coffee, and nice pastries.

(SC) Hello this is a question for Caroline really, Susan Corrigall, Copyright Officer at the [National Records of Scotland](#). I must say that I was very, very impressed at your hit rate at tracking down the copyright holders, eighty plus percent. What I would be interested in is that you spoke about the Wellcome had worked out the list of sources which they felt constituted a diligent search, is it possible at all to give any indication of the time which it took to trace the copyright holders, to do your reverse genealogy?

(CHer) Short answer not really because it did vary a lot. The sources were things like Who's Who and then if we thought there was a literary

estate then the [WATCH File](#) which people may or may not be familiar with which lists contacts there. I feel slightly bad saying this as a trained librarian but things like [Google](#) and [Wikipedia](#) with the obvious caveat that, that wasn't the only source but it was things like that where you'd maybe build up a picture. If Wikipedia tells you that these are the names of the children then you search the names of the children in Who's Who or something and see if it tells you that their father was the person that you think it was and put it all together. Because obviously if you get nothing back then it doesn't take long but I honestly, I mean it could have been an hour or two maybe for those that were more difficult to find or when you weren't getting anything back, sorry that's a very long way of saying I'm not sure I can really put a figure on it because some were very quick and some took a lot longer.

(SC) How long were you working on finding the copyright if you could say, was it a two year project how many weeks, months?

(CHen) We didn't work intensively in copyright the entire time though.

(CHer) No, I was juggling it with other things, so.

(CHen) And also we had a big push to find a lot of people, send off a bunch of letters, and then wait and do nothing. And then things come back and they suddenly start going oh can we have copies or our fifty letters and so that added up time.

(CHer) I can tell you that if someone's got a list of about twenty files that there are letters in and you end up with the copies about an inch thick going in to the envelope, that takes at least a day, maybe more.

(RD) I think one of the things to appreciate about Codebreakers is it was never a project that was designed to study rights clearance - that was not part of the brief at the start. It was when we came in and said oh, you're doing amazing things here, let's think about the implications for rights clearance and times and costs and making life easier but that again we'll talk about next steps at the end. But I think anybody, any archivist in the room who's contemplating any kind of digitisation project that is one of the things that you might think about tracking, because that's really useful information, really useful evidence that

helps us begin to think about what copyright policy should be in this space. So if there are fudgy answers here it's because they just weren't thinking about this stuff at the start.

(CHen) And also we're kind of pre-limiting ourselves to those who are already easier to find. If we'd suddenly opened up the floodgates and said we're going to trace every single John Smith...

(LR) It makes your job impossible.

(CHen) It would be infinite.

(RD) Okay thank you very much for your questions. Let's take a break, we reconvene at quarter past. We'll have a second panel, which is Victoria providing some reflections on the Codebreakers project and then a general Q and A about the nature of risk and what are appropriate risk criteria for archivists to think about when you're doing something like this. Thank you very much.

PANEL 2

ARCHIVES, DIGITISATION AND MANAGING RISK

Speakers

Victoria Stobo (Research Assistant and Postgraduate Researcher, CREATe, University of Glasgow) – hereinafter **(VS)**

Chair: **David Mander** (Vice-Chair, Board of the Archives and Records Association, UK) – hereinafter **(DM)**

Questions and Answers

Christy Henshaw (Digitisation Programme Manager at the Wellcome Library) – hereinafter **(CHen)**

Caroline Herbert (former Content and Metadata Officer at the Wellcome Library) – hereinafter **(CHer)**

Lesley Richmond (University Archivist and Deputy Director, University of Glasgow Library) – hereinafter **(LR)**

Ronan Deazley (Professor of Copyright Law, University of Glasgow) – hereinafter **(RD)**

Chris Bird (Senior Legal Counsel at the Wellcome Trust) – hereinafter **(CB)**

Susan Corrigall (Copyright Officer at the National Records of Scotland) – hereinafter **(SC)**

Cathy Williams (Head of Collections Knowledge, TNA) – hereinafter **(CW)**

Judith Etherton (Freelance Archivist) – hereinafter **(JE)**

Tim Padfield (International Council on Archives) – hereinafter **(TP)**

Natalie Adams (Information Services Manager, Churchill Archives Centre) – hereinafter **(NA)**

Craig Alexander Moore (Records Manager, the Bar Council) – hereinafter **(CAM)**

Ellie Robinson (Archivist, LSE) – hereinafter **(ER)**

Ian Anderson (Director of Museum Studies, HATII, University of Glasgow) – hereinafter **(IA)**

(DM) I'm David Mander and I was very pleased to be invited to this session in my capacity of Vice Chair of the [Archives and Records Association](). One of our Core Courses is on copyright; it's an area of considerable concern to our profession. In my other capacity in my working life, I'm a partner in [Creative Cultures](), an arts archive and heritage consultancy. I suppose my qualification for chairing this session is that I have over thirty years of experience as a custodian, and subsequently as a consultant, grappling with rights for publication, rights for digitisation for a batch of projects, and then having the temerity to advise clients on what they should do with digitisation and evaluating the outcomes afterwards.

Without further ado, we have a slight change to the programme. Sadly Simon Chaplin who was going to be the panel respondent and has appeared on your programme can't be here today for family reasons, so Victoria will be our sole speaker. At the end of the session our three previous panel speakers will be coming up to join Victoria for a proper, full, question and answer session.

So to introduce Victoria, she is currently a Research Assistant at [CREATe]() in the [School of Law]() at Glasgow University. The majority of her archive training took place at the University of Glasgow Archive Service and she gained an [MSc in Information Management and Preservation]() from the University of Glasgow last year. She went on to work as Research Assistant on the [Copyright and Risk: Scoping the Wellcome Digital Library project](). Victoria tells me that this is the first week of her [PhD](), so all of us in the room are all her potential subjects, you are hereby warned. Victoria will now speak to us on copyright and risk, scoping Codebreakers, and the Wellcome Digital Library.

(VS) Thanks for that introduction David. First I should explain my involvement in the Copyright and Risk project. Since February I've been working as a part time Research Assistant in CREATe which is the Research Council Centre for Copyright and New Business Models in the Creative Economy and that's based at the [University of Glasgow]().

Copyright and Risk aimed to assess the merits of, and problems encountered by using a risk managed approach to copyright clearance

for archival digitisation projects, and the Wellcome Library Codebreakers project was selected as an ideal case study because they used a risk management strategy, because of the scale of the digitisation involved, and because they hold both library and archive collections. Copyright and Risk has three deliverables: a working paper, a project report and this conference today. The working paper, [Archives and Copyright - Risk and Reform](#) which I co-authored with Ronan was published earlier this year. It reviews current and proposed changes to UK copyright law and specifically the exceptions provided for archives and libraries. This was further extended by a review of available literature on the digitisation of archival and library collections for publication online and we contextualise that with specific examples.

Existing literature reveals that rights clearance procedures impose prohibitive burdens on cultural institutions through the cost of staff time both in training and diligent search and the process of contacting rights holders. It also indicates that in most cases the results of rights clearance processes are unsatisfactory. Either copyright holders cannot be located and traced or those who are contacted don't respond to permission requests.

Archives have the added complication of dealing with larger collections of material, material which is often unpublished at the point of deposit, and typically includes higher proportions of orphan works when compared with library collections. Archives also differ from libraries in that most of the material in archive collections was collected for non-commercial purposes and will inevitably contain far more sensitive data.

Documented examples of archival rights clearance projects are scarce and the studies which do exist tend to be library focused and deal with published material. One or two fairly large scale studies have been conducted, but in general there's a lack of evidence concerning rights clearance in archival digitisation and especially in relation to projects that employ risk management strategies in a sector which is known to be highly risk averse.

In addition to the working paper I've spent time this year between April and June interviewing staff at the Wellcome Library and also at the third party archives that were involved in delivering the Codebreakers project. I've also had access to some of the documents and policies created during the project. I'm using this material to prepare a [project report](#) which should be available later this year but I thought I'd share some of my findings with you today.

We heard earlier from Caroline and Christy about the results of the project and before I begin I should emphasise that I'm commenting on the risk management strategy used for the archive digitisation and not the library digitisation. I should reiterate from earlier that Wellcome took the view from the beginning of the project that strict compliance with copyright law would be impossible because rights couldn't be cleared in everything.

The results show that the success for finding contact details was relatively high at eighty four percent, but given that the total number of rights holders was controlled this isn't really surprising. Other projects that have comprehensive rights clearance have typically yielded a far lower return. For example when the Wellcome Library previously attempted to identify and contact the rights holders in a collection of [AIDS posters](#), contact details could only be found for half of the rights holders. Of the permissions sent, seventy nine percent of those contacted replied with ninety eight percent granting permission and for only two percent was permission to digitise material expressly refused. In one of those instances the rights holder was more concerned about sensitive information contained in the material and not the intellectual property per se. A substantial number of rights holders have simply not responded at all. Non-response to request to digitise material is a real problem especially if we consider that without permission most archivists are unwilling to publish material online.

Wellcome dealt with non-response by reviewing and reassessing their risk criteria, so where non-respondents were known to the Wellcome then the digitised material was made available. For non-respondents that were not known to the Wellcome but were nevertheless deemed to be low risk, for example scientists and politicians whose work was

created in a non-commercial capacity, their work was put online after a further two month hiatus. For non-respondents who were deemed to be high risk, the decision was taken not to post anything online without securing express permission first.

The sixteen percent of rights holders whose contact details could not be found can be classified as orphan works. Orphan works continue to be problematic for archives. It's been estimated that between twenty one and thirty percent of individual archive collections are made up of orphan works. The Wellcome have made these orphan works available online in batches, but given the proposed licensing regime to regulate the use of orphan works, it may be the case that making these works available in this way will become untenable; depending on the numbers of orphan works in the collection and the amount of budget an institution can afford to allocate to licensing fees. We'll hear more about the legislative proposals in the third panel today.

The Wellcome's risk management strategy has two broad aims: to manage the risks associated with publishing copyright material online, and to manage the risks associated with publishing potentially sensitive material online. They managed the publication of copyright material using risk criteria to identify rights holders who were likely to object to publication, and they also had a due diligence standard for locating and contacting those rights holders. They managed the publication of potentially sensitive material using their Access to Archives policy, which provides guidance that can be followed for all physical, digitised, and born-digital material which is held across the Wellcome institution, and that takes into account both search room and online access to sensitive personal data. In addition to this they have an established take-down policy which pre-dates Codebreakers and applies to all the material available on their website.

The risk criteria consisted of low, medium and high risk, and low risk was a default category which everything not considered medium or high risk fell into, and this is the criteria for medium risk. The Wellcome found that once those criteria had been circulated and the third parties asked to compile lists of medium and high risk rights holders, the lists that they received back from this process were fairly

long. The complete list once it had been cross-checked ran to several hundred names and through an iterative, negotiated process between Wellcome and the third parties it was reduced by nearly two thirds and this suggests at first the third parties struggled to apply the risk criteria in the way that the Wellcome initially anticipated. The third parties also wanted to modify the criteria. Some of the third parties felt that the emphasis was on commercial authors who would sue whereas they were concerned about maintaining good relationships with their depositors and stakeholders. For this reason the third parties asked that specific rights holders who were not identified by the Wellcome criteria should be contacted as a courtesy, and in some cases they also identified stakeholders who were not directly involved in Codebreakers but who they felt should also be contacted as a courtesy. The reason given for this was to maintain those good working relationships with stakeholders and depositors.

One of the first things I noticed in the interviews was that some of the third parties felt that the risk criteria had been written with library and published material in mind, and I think that judging the first of these medium risk criteria requires specific subject knowledge. The second queries the use of a resource which was developed for chasing the rights holders of published material. And the third criteria also leans towards published material.

If we look at the high risk criteria a few questions spring to mind. Again, with the first criteria subject knowledge is called for. Which sources should an archivist consult to gauge whether someone actively defends their copyright? The third criterion plays directly to the concerns of the third parties who wanted to maintain good relationships with their depositors, and the fourth criterion is unusual in that if someone is determined to defend their copyright it doesn't matter if you publish a hundred of their letters or just one, they're going to complain.[1]

_____

[1] In practice, the Wellcome Library did not rely on the fourth risk criterion when assessing collections for medium and high risk rightsholders.

In the same way that third parties modified the risk criteria, I'd like to ask all of you in the Q and A session and throughout the rest of today to reflect on these criteria in light of projects that you've worked on in the past, or that you're maybe going to work on in the future and consider which of the risk criteria could work for you and which cannot. So, if you're a small institution, do you have the staff with the subject knowledge required to make these kinds of judgements? And if you don't, do you have contacts or networks that you can rely on to get access to that kind of knowledge? Do you already adopt specific risk based criteria that are similar or different to these? And do you strongly agree or strongly disagree with anything that's listed here?

The Wellcome Library use a percentage checking strategy to identify sensitive data in their collections and this is outlined in their Access to Archives policy. Given the scale of Codebreakers, this was the method they used to identify and restrict access to sensitive material online. To reiterate what Caroline and Lesley were saying earlier, sensitive data in archive collections selected for digitisation is important because checking collections for that data prior to publication is such a labour intensive, time-consuming exercise and that's without even having to consider rights clearance on top. It's also, and I know this is developing into a theme, a consideration that library mass digitisation projects don't have to account for.

Each of the staff members I interviewed agreed that identifying sensitive data in the collections was of greater importance than identifying and contacting rights holders and this is borne out by the fact that the Wellcome managed to convince the five partners that they worked with, some of whom in previous projects had engaged in comprehensive rights clearance to take a risk managed approach to copyright compliance. They did this by offering them a contract in which the Wellcome assumed responsibility for publication, but what they couldn't convince three of the partner archives to do was to digitise any material that might contain sensitive information. I think this again reinforces that managing sensitivity is of greater concern to archivists than copyright compliance but also that the acceptance and

management of divergent practices and expectations was an important contributing factor to the success of the project.

I should also note that the take-down policy was developed as a result of the AIDS posters project which was the Wellcome's first attempt at a comprehensive rights clearance exercise, and the take down policy is useful in terms of risk mitigation in that it applies even to material for which permission to publish has been given in the past but is then subsequently removed.

I'd like to say again that I think Codebreakers has been a tremendous success both for the Wellcome Library and their partners, both the digitisation effort and the rights clearance effort. The Wellcome have received only one take-down request in relation to Codebreakers and this appears to have been precipitated by concerns about the sensitivity of the material rather than their intellectual property. There are now two million images available on the Wellcome Library site with more to be released and future digitisation projects to start.

I think that consistency of communication was an important factor in the success of Codebreakers. This extends from the formulation of policies and best practices through the letters and licence agreements sent to the rights holders to the information available on the Wellcome Library website.

There were two aspects of the project that Wellcome emphasised throughout their communications. The first was that the project was non-commercial; material was being made available to researchers and the public in the public interest. No content would be available for commercial licensing without permission and no fees would be paid to archives rights holders. And the second was the explanation of the take-down and Access to Archives policies to reassure rights holders that the material would be checked for sensitive data and that it could be removed from the site at any time. Rights holders were also reassured by the fact that readers would be required to register to use the Wellcome Library digital player and that the Wellcome's Terms of Use would have to be agreed to.

This is especially relevant given that any objections to publication have been around sensitive data, and throughout the Copyright and Risk project the difficulty of defining and managing sensitive data within the collections, has been a counterpoint to the focus on increased access through risk managed rights clearance. I think more research on access policies and sensitive data in archives needs to take place. Communication between the Wellcome and the third party archives was also a contributing factor to the success of Codebreakers, and from the interviews the impression given was that Wellcome was very willing to listen to and learn from the third parties, and vice versa.

However, some of the third party archives will continue to follow strict compliance with copyright law, at least when digitising smaller collections of material. At least two of the third parties said that collections which did not involve complex rights clearance would take priority in terms of digitisation, a decision which is not surprising but is itself a form of risk deferral and this is linked to the concerns around reputational damage mentioned earlier. Several of the third parties and the Wellcome itself stated that reputational damage was the main concern during Codebreakers. This is also borne out by the fact that the third parties wanted to ask permission of rights holders not identified by the Wellcome risk criteria and in some cases contacted individuals purely as a courtesy.

Damaging your reputation as a trusted and reliable repository is indeed a serious risk, if you consider that the reputation of the Wellcome Trust and by extension the Library, was important in securing the participation of rights holders in Codebreakers in the first place.

Archivists at one of the third party archives also pointed out that courtesy contact is useful in terms of generating enthusiasm for the project, encouraging future donations and instigating outreach activities.

Taken together this shows that working with the Wellcome on this project has not significantly changed the appetite for risk of the third parties, but I think this is also the result of the fact that there are different levels of experience with rights clearance and rights related

risk management at each archive, and that one project will not be enough to change these attitudes.

I think that Codebreakers worked successfully because the Wellcome and the third parties were willing to accept divergence of working practices and expectations, the results of the rights clearance process they have engaged in have been better than many other comprehensive clearance exercises, and they've only received one take-down request so far.

I think the Wellcome have demonstrated considerable and unusual appetite for risk, and they may be able to do this because they have the resources, but there are elements of their approach which, I hope you agree, can be applied by the rest of the sector. The Wellcome strategy could be relevant if you do not have the staff, time or resources to engage in comprehensive clearance and if you're willing to take on a certain amount of risk, with the proviso that even by reducing the number of rights holders contacted this was still a lengthy, complicated process for the Wellcome.

Individual archives will have to define their own appetite for risk based on the type of collection you want to digitise, the level of access you intend to provide, the effect of copyright and data protection law and the type of digitisation that you intend to engage in. We'll be hearing from Peter Hirtle later today, but I think his advice in the book ['Copyright in Cultural Institutions'](#) is relevant especially in light of Codebreakers. *Explain* your project and what you're doing clearly, *solicit* information where possible, *document* your due diligence and *avoid commercial use.*[2]

Before I finish, I would like to repeat my questions from earlier about the risk criteria. If you're a small institution do you have the staff with the subject knowledge to make those kinds of judgements? If you don't,

---

[2] Understandably, some archives may be looking to engage in commercialisation activities, and avoiding commercial use may not be a realistic aim. In these circumstances, taking a risk-managed approach to rights clearance could be inappropriate: if there is an expectation that profit will be made from digitisation, it may be appropriate to obtain express permission from rightsholders.

do you have access to contacts and networks that can get access to that knowledge for you? Do you already adopt specific risk based criteria that are similar or different to these? And do you strongly disagree or agree with any of these criteria?

These are my final points: the majority of rights holders have said yes, they don't want money and if you're prepared to accept that risk you can make more material available online. If you're engaging in digitisation, you're thinking about it or you have previous experience it would be really useful if I could speak to you over the next year or two. And as you can see we've got a film crew and photographers in, and the resource that we're preparing for this event will hopefully be ready by the end of this year. Once it's finished you'll find it at the event page on the CREATe website, and I've also included a link to the working paper that was published earlier this year.

I thought I'd finish just by putting the risk criteria back up and for the Q and A, and we can maybe talk some more about what your opinions of the risk criteria is and what your own risk criteria might look like.

(DM) Victoria, thank you very much.

[Applause]

(DM) Can I now ask our first session panel, Christy, Caroline and Lesley to resume the stage and we can then open up for questions. The floor is yours and roving mics are coming round. Again, please say who you are and your institution before asking the question.

(CW) Cathy Williams from [the National Archives](#). I've got a couple of comments to make. Firstly, the issue there about avoiding commercialisation: some archives need to think about commercialising what they hold, simply to sustain themselves. For some archives, their model of sustainability in the future will include commercial use of what they hold, so I think that's something to consider.

The other issue this morning Lesley was talking about, yes, archives are the result of human activity but for many archives even in what might seem quite mundane collections you could trip over material that was

produced for commercial use. Don't forget, there are many film archives, there are photographic archives, the National Archives, even when we digitised for a London 2012 project, even hidden among Crown copyright material from government, are cuttings from newspapers. It took us a very long time to trace owners of old sporting magazines and so on, and that's hiding in amongst official public records. It is complex and I don't want to over complicate things. One of the things we're now embarking on is a programme called [Archiving the Arts](#) and we would welcome working with CREATe on aspects of artistic archives, where some of the material is about an artist's bread and butter and whereas we might want to hold it and make it publicly accessible, but when we're looking at deposit agreements, copyright agreements, it gets very complicated

(DM) Thank you Cathy, who wants to comment or respond from the panel.

(CHen) I'll just comment on the idea that people amass a lot of different things over their life, it gets into their archive, and some of that would have been newspaper clippings, magazines. You do need to know what's in your archive before you make any kind of blanket assessment about what kind of copyright clearance you're going to do, because you've got assess those things as well. From our point of view, we don't really think that a stray newspaper clipping is a very high risk at all. An entire magazine is more of a high risk. We may not clear it, we may not even digitise it, because you can probably get it somewhere else. So yeah, absolutely it's not just all things people have written on napkins.

(DM) Yes, Judith first and then Tim.

(JE) Judith Etherton, I'm a freelance archivist and my question is about the commercial aspect as well, it follows on from something that Lesley said at the end of her talk about how your Codebreakers project was non-commercial and this was the copyright clearance you did. But now the images are out on the web, people will want to use them, probably, commercially and how are you going to deal with this going forward. I would like to actually have a little bit more discussion on this aspect.

(CHen) It hasn't happened yet, but we would have to look at each request individually and if they were asking for commercial use of something where we don't know the contact details, because it's not someone we've traced before, we would ask them to trace the copyright holder and get permission to use it. We don't have a service that provides that for people because they want to use an image. We don't have the resources to go out and trace the copyright for them, so if they really want to use it for commercial purposes, they've got to do that, then they can tell us that they've got that permission or they can show us that they've got that permission.

There might be things that turn out to be out of copyright, but we never had a chance to really look at it when we were doing a mass digitisation project. If somebody says, "This one image, I'd really like," we look at it and that one happens to be out of copyright, you never know. You've got to take it on a case by case basis. If we did have the details of the people, and this does come back to Lesley's point about keeping things up to date, and we felt that it was something we wanted to do, we could potentially contact those people for them on their behalf and see if they would be happy. Commercial use is a bit grey sometimes. It could be something that's a really valuable use, and it could be very academic but it's also commercial, and people might be perfectly happy about that. But at the moment we don't have contributor agreements with people where we pay royalties. We don't have any of that in place for this content.

(DM) Does any other panel member want to come in?

(CHer) I was only going to say, I think in effect what Christy's describing would be much the same for any archives, with users sitting in the reading room, looking through a file and thinking, I want to put this letter in my next book. In some ways the fact that the material's online wouldn't really affect that part of how the library would respond to a request.

(DM) Thanks, Tim you wanted to come in?

(TP) Yes, Tim Padfield. I'm retired but I'm down on the programme as representing the [International Council on Archives](#). I have one comment and one question. The comment is, I recognise the need to accept risk but personally I would urge archivists to make sure that they get approval from their senior management, rather than take any risk themselves. Always get senior management agreement so that they take the risk. My question is one about the rights owners. The assumption seems to be that you do all this diligent research and you come across a single rights owner. My experience with wills is that most people when they leave a will do not mention copyright at all, and so copyright would be among the residuary beneficiaries' benefits, and there might well be several of them. If I leave everything I own to all my children, I might have four children and that means there will be four owners that you have to trace. I think the assumption that there is a single owner, even through several generations, is probably making it sound even easier than it really is.

(DM) Who wants to pick that one up?

(LR) Well, I would agree. One of our three collections that we were digitising, there were two marriages with five children, and both wives were still alive. I was going to ask people, do you think you have the copyright? Were you still on speaking terms with your first husband? It was extremely difficult and Glasgow University Archives could not crack that, we tried and we tried and we tried. It was Wellcome who managed to work out where the first wife or the second wife was living. It is split between all the family, but what Victoria said earlier has now become very good, because we now have different members of the family very enthusiastic about Codebreakers. They've been to every event that Wellcome's had about Codebreakers, different members of them, but I don't know if they can all get into the same room at the same time and not have a fight.

(CHer) Can I just come back? Yes, I agree we were making assumptions about copyright, next of kin and residual estates, so yes I suppose there is an argument that we should have checked every will and gone through all the family. In practice, we found that because this was a limited list and we were looking at people with high public profiles,

generally if we did contact one of the children, there was someone who was more used to speaking on behalf of the family to answer such requests. It was only in some cases where they would bring other siblings into the loop, and say, "Check with my brother or my sister as well." If we tracked one descendant, we were relying on them to do the decent thing and share with us if there were others who should be involved, which I think generally they did.

(NA) I'm Natalie Adams from Churchill Archive Centre and what I wanted to say is more of a comment. I've spent a long time working in the publication of the Churchill papers which has been a commercial project, so it's different from the Wellcome project in that key sense, and we also had to be more comprehensive in our clearance. But what we found at the end shook down to a very similar situation that the Wellcome have found. Most of the people said yes, a few more of them than the Wellcome had asked for money, but that's completely fair enough as it was a commercial project. But what we also found was this feel good thing where people were amazed, "My grandfather wrote to Churchill, can I see it? You want to publish it? How brilliant, yes." Don't necessarily think that because it's a commercial project that's going to completely change the landscape that the Wellcome have found. because I don't think it necessarily does. Also, I wanted to echo the very difficult decisions about which rightsholders to contact, we did quite a lot of clearing all the people we thought it might be, rather than trying to decide who we thought it was.

(SC) Susan Corrigall again. I was intrigued by Victoria's estimate of the quantity of orphan works in archival collections. You estimated between twenty one and twenty five percent of our archival collections would be orphan works. It strikes me that that's rather optimistic thinking about collections in my own institution, I would completely flip that percentage it would be much more like eighty percent, ninety percent. But I can see that in other institutions, other collections it would be quite different. I think that would be a highly variable figure depending on the nature of the institution and the nature of the collections held.

(VS) Absolutely, I would agree. That figure's based on the JISC *In From the Cold* study of orphan works and that was a Europe-wide study. I think that's one of the problems with the evidence that is available, it's conducted at a very high level so when you do try to apply it in specific instances, it doesn't appear to match up or it doesn't seem quite as relevant. That's an area where more research has to happen and it could happen in Britain specifically, not just at a Europe-wide level. I would agree that it does depend on the collection and the institution in question.

(DM) Craig next, at the back.

(CAM) Craig Alexander Moore of the Bar Council. I'm interested in the role of lawyers in risk taking, because as I'm sure many of you are aware, lawyers are some of the most risk averse creatures on the planet. There have been many compliance issues where I've put together the presentation, I've vigorously argued at Board level, everyone nods and says yes that all sounds very good, let's just run it past legal and I've never heard about it again. If a solicitor is going to act as the final arbitrator on risk, how can we as practitioners address that?

(DM) Who wants to tackle that one?

(CB) Chris Bird, Senior Legal Counsel at the Wellcome. I'm not sure what I can say about this, but one thing I would say is we have various levels of lawyer in this organisation, I have a boss whose job title is the General Counsel of the organisation, and we face some issues getting it through the legal hurdles and we have lots of debates. As I said to some people in the coffee break, it quite often ended up with me banging the table and saying, "Look, we are doing this, get comfortable, let me help you get comfortable but we are going to do it." And if we're not going to do it, then we really need to think about how we act as business partners to the organisation. It does need one of the lawyers at least to be quite integrated into what the business wants to achieve, and to have quite a developed view for themselves of what risk really looks like, and in that case it was me.

(DM) Can I endorse that. I think there is a strong risk in a large organisation of the legal team having insufficient expertise and you can then present a case that can either then be met with that awful thing, a long silence, or even worse, wrong information coming back to you. With due respect to my learned friends, they're not always the fountain of information that you feel they ought to be. Key communication is needed here, otherwise what you get is a rushed judgement which won't help you, and I'm not going to name the organisation concerned, but can actually be plain wrong and circulated and unhelpful if this is a partnership project.

Any more questions?

(ER) I'm Ellie Robinson from the [London School of Economics](). I have two questions, if that's all right. Just quickly for Christy, when you were managing the digitisation for Codebreakers, was the digitisation run concurrently with the copyright clearance work, or did you wait for permission until you started digitisation?

(CHen) It depends on the project, but with archives we started first and then we sought copyright clearance. We weren't putting it online at the time, we were just digitising. We didn't put anything online until the copyright clearance procedure had gone through, but that ended up working out quite well in a practical sense because so many people did ask for copies and we had them for the most part, once we established how exactly we were going to do it. Now when we do new collections we start copyright clearance as soon as we know what collection we're doing.

(ER) And for the metadata for the digitised objects, did you go back and update the rights information for those?

(CHen) Yes, so there's various different ways we've updated the rights, we've also assigned licenses to material. If we need to, we can assign different types of licences to material. That came in with the books where copyright holders had a lot of different choices of access, whereas with the archives we really just wanted attribution non-commercial access. For the most part we didn't have to do too much of

that. If somebody said they wanted to be credited in a certain way, we can put that back in the catalogue record but that does add on extra effort, it's just another thing we couldn't do if we were trying to clear every single copyright holder.

(ER) And if I can ask Victoria, with the risk criteria that you've been examining, I was wondering if the size of the institution has any impact on that? Obviously the Wellcome is a very high profile, well known institution with a large audience, and I was wondering if smaller institutions with a smaller audience and smaller number of people who would ever potentially see the material that you're putting online, whether that has an impact on the level of risk?

(VS) Yes, I think it will. The risk criteria that the Wellcome used, has been the only risk criteria that I've looked at. Over the course of my PhD I would like to get in touch with smaller institutions that are making material available online and look at what kind of risk analysis they do before they do that, so I don't really have very much information on that at the moment, but hopefully I will be able to generate that information. I think it will have a definite effect on their risk analysis.

(IA) Ian Anderson, University of Glasgow. Question for Lesley, you mentioned how glad you were that Wellcome was accepting the liability in terms of this project. I'm just wondering whether having had the experience of it, you would be more willing to accept liability? Or is it preferable, particularly in collaborative digitisation projects, for one of the partners to accept liability on behalf of everybody, perhaps the partner with the greatest appetite for risk?

(LR) Part of being grateful for Wellcome being the publisher, so therefore taking on the liability, was really because this project was going to be such a high profile project, and we were dealing with twentieth century material where there's definitely going to be higher risk of copyright holders perhaps wanting the material taken down etc. I do have a large appetite for risk generally, although when Victoria was talking about organisations that look at their lower risk material first, we're probably that at the same time. I've got a large appetite but I'm

still cautious. We just have such a lot of material that is out of copyright that we can digitise so, do that first, but at the same time take part in projects like this, where you're learning how to do the recent material that's more risky. Again, it's the nature of the collaborative partnership that you're in, but if you're in collaboration with a national organisation and you can persuade them that they'll take on the burden of doing the due diligence over the copyright holders, then there is an advantage for one taking the lead overall on the copyright, and actually being the publishers. But it really comes down to what that body's putting into the whole project, and Wellcome were taking the lead on so much and we were just the partners that joined in. If I was going to go into partnership with other Scottish universities, for instance, I don't see that one of us probably wouldn't take that lead, unless we'd got money from a funding council, so it's where your funding is coming from.

(CHen) Do you mind if I just add to that, because we were taking responsibility for the risk of making material available online through our own website, but for example Coldspring Harbour Laboratory have also made their Watson and Brenner material available on their own website. If they didn't think that the copyright clearance procedure was good enough, they would have had to go and do more themselves in order to fulfil their own appetite for risk for their own website, we wouldn't have covered them. We weren't taking the risk for them doing stuff with their images, if that makes sense. It sounded a little bit like you might have thought that we were taking on all the publishing, all the risk, that isn't necessarily the case.

(LR) It seems completely nonsensical to have this material available on two websites. It may lead researchers to think there's more material out there, so we're not publishing the material that we've digitised online ourselves. We've taken part in the Codebreakers project, so I think that would also make a difference too. You could see projects where it wasn't being published in one place, and each institution was publishing themselves.

(DM) I'm going to exercise chair's privilege and squeeze a question in of my own here. Victoria, but for other panel members as well, you talked about the very high costs in staff time of identifying risk. Do you

think you, or other panel members, can come up with formulae that will help people who are putting projects together for grant assessment, to say here is a methodology for you to cost this out and go to the funder?

(LR) Yes, I think it will be possible.

(DM) Because that will be really useful actually, something that quantifies, not says this is the flat figure, but here are the stages, here are the things you need to think about in order to arrive at your costs with all the risks that we're all very well aware of. That would be very helpful. Thank you for a very lively question session.

PANEL 3

ARCHIVES AND COPYRIGHT:

THE VIEW FROM THE UK

Speakers
Robin Stout (Policy Advisor, IPO) – hereinafter **(RS)**
Nick Munn (Deputy Director for Copyright, IPO) – hereinafter **(NM)**
Tim Padfield (International Council on Archives) – hereinafter **(TP)**
Chair: **Martin Kretschmer** (Director of CREATe, University of Glasgow) – hereinafter **(MK)**

Questions and Answers
Ronan Deazley (Professor of Copyright Law, University of Glasgow) – hereinafter **(RD)**
Susan Corrigall (Copyright Officer at the National Records of Scotland) – hereinafter **(SC)**
Cathy Williams (Head of Collections Knowledge, TNA) – hereinafter **(CW)**
Ellie Robinson (Archivist, LSE) – hereinafter **(ER)**
Anna Vernon (Licensing and Copyright Assurance Manager, British Library) – hereinafter **(AV)**
Paul Gibbons (Information Compliance Manager, SOAS) – hereinafter **(PG)**
Pam Temple (Community Archivist) – hereinafter **(PT)**

(MK) Good afternoon, I'm Martin Kretschmer, I'm chairing the next session. I'm the Director of CREATe, and I think today is an event of the kind CREATe will be very keen to advance over the next years. I think a key aspect of this is that normally policy makers hear about the issues

in their field through secondary sources. They will have somebody speaking on behalf. There will be a trade association, there will be somebody who's a paid lobbyist and they will get a filtered story which suits particular parties, and one of our aims as a centre is to break this. We want the voices of the primary users, right holders and creators to be part of the process, and we want to move that beyond the anecdotal. So, in some sense, what we have in the very first session this morning was a narrative about rights clearance in one particular setting, and that may be a story, may be something that feeds into policy making and may be used, but the key is to turn that into something which can count as evidence, and which can be put to the test of a process of scrutiny, and this event is part of this. It's a very important part of the way in which we methodologically want to advance this area.

So thanks very much in the first place to Ronan, whose idea that whole process was, and to Victoria, who condensed and systematised the anecdotal stories into something which can then be rolled out across the sector.

My second intervention is really a provocation. Under UK damage jurisprudence, which goes back centuries, common law principles, the damages you would be awarded will put you in as good a position, as if no wrong had occurred. That's what the claimant gets. As if no wrong had occurred. Therefore, I would submit that under UK law as it currently stands, the financial risk for UK materials is pretty much zero. If we see the process evidence this morning by Wellcome, there is little risk that the additional damages available under [Section 97 of the Copyright, Design and Patents Act](#) would come into play. There's clearly no flagrancy involved at all. It looks to me what we really need from the archive sector is a process, and a willingness, to be taking risks in the right manner, and we may not need intervention by the lawmakers. Okay, so is that true or not? It's a thought which you might have.

So, I ask the first two speakers to introduce themselves. I know them both personally from my secondment, so it's better if they introduce themselves.

(RS) My name's Robin Stout and I'm a Policy Advisor at the [Intellectual Property Office](), working in our copyright team. I have a long title, I was told not to give my long, complicated, civil service title, so that's who I am today. My job here today is to talk about the changes that we're going to be making to copyright law, hopefully very soon, around copyright exceptions. What's a copyright exception? A Copyright exception, otherwise known as a permitted act in the law, is basically something you can do with a copyright work without having to ask someone first. We're going to be updating the ones that already exist, and we're going to be introducing a few more, and I'm going to take you through some of the consequences of that.

To begin with, a bit of background. In 2010, the Prime Minister asked Professor Ian Hargreaves to have a look at the intellectual property framework and see if it was supporting innovation and economic growth. He did this and he made a number of recommendations on all sorts of areas of IP, including copyright and in particular in relation to exceptions. The conclusion he reached was that copyright's very important and it's important because it encourages creativity, and it rewards creativity and creators, and it helps to support the creative industries. But sometimes it gets in the way of people doing useful things and unnecessarily so, that the harm that's done by those activities is very small, and the benefits could be considerable. His argument was that, as it says here, the UK should introduce copyright exceptions into its law, to the maximum extent that it can do in the European framework. The government accepted that point and in typical government fashion, decided to go and consult for a year to work out what it should do, and it's quite a huge process. The [Copyright Directive]() allows us to have up to twenty types of exception, we consulted on most of those types and asked people whether they thought it was a good idea or not, and the basic starting point was that people had to argue why we shouldn't do it, they had to make that case. After this process, we received about 450 consultation responses, we wrote 9 [impact assessments]() and we published a document called [Modernising Copyright](), which said what we're going to do. The upshot is that we're going to make amendments to a whole lot of existing copyright exceptions, and we're going to also introduce 3 new ones:

one for private copying, one for parody and one for text and data mining.

As I said, that process began with the review starting in 2010. It's now 2013 and where are we now? We are currently preparing the legislation and we're hoping that that will come into force next year on the 6th April. We're currently going through what we're calling our technical review process, where we've published the regulations, we're asking people for their views on them and to make sure we've got it right. We've had some public meetings and not quite the same number of responses as our policy consultation but certainly, just in relation to the first set of 4 exceptions, we had over 100 responses. This is people commenting on quite technical law, so it's an area where there's a surprising amount of interest.

The goal of the legislation is to implement the changes set out in the document we published last year, and the basic starting point for it is that the legislation should deliver that. Our consultation, our technical review of it, is to make sure it does that effectively. All sorts of people write in to us and complain and say, "You should have done this, you should be doing that, you shouldn't be doing this," but we just want to get it implemented and get it implemented well.

One thing we want to do, the overriding purpose of this, is to bring it all up to date and make sure copyright doesn't stand in the way of uses that don't really cause any harm or undermine copyright, but could provide a lot of benefits to people. But also, going back to Martin's point, one of these is to do with format shifting, CDs and other material. Copying a CD to an MP3 player in the UK is still technically illegal, and so we'll be introducing an exception to say that, if you've bought the CD and you're just copying it to your own device, then that would be allowed. The idea is that that will apply to other materials as well, so the same would go for an ebook or other material. People say to us, "We're all doing this anyway and no one's ever complained," but I think there's an important point; the law should be sensible and people shouldn't have to break it to do reasonable things. Very often, behind the scenes, there are things that people don't see are going on, so there are often hidden charges in different places, licencing charges. People

might not go after an individual but they might go after an organisation that that individual works for, and so there are a range of reasons why these are justified.

So, when we go about preparing this legislation, one thing that we're keen to do is to simplify some of the law. If any of you are ever bored enough to want to read the [Copyright Designs and Patents Act 1988](#) as amended, then you'll see that it's fabulously complicated and very difficult, even for people working in the field to understand. One thing we're trying to do, when we bring in these amendments, as well as implementing the policy, is to try to simplify it a bit along the way. One way we'll be doing this, in particular relation to libraries and archives, is to get rid of some of the supporting legislation, and in particular the [Copyright (Librarians and Archivists) (Copying of Copyright Material) Regulations](#), which is an extraordinary piece of work. You have a lot of specific provisions in the Copyright Act saying that libraries and archives can do this and that, and I'll tell you in a moment how we're going to update those, don't worry, I'll get there. But underneath these, you have all of these provisions, you read them and you think, "I can do this," and then you realise a lot of it is defined in these other regulations, which includes a definition of what a library is. Then you think, "I need to look at the Act and then I need these regulations to find out what a library is, am I actually a library?" Then you go to the Schedule in the Regulations which sets out 6 classes of library, and the final class is more or less 'Any Other Library'. An archive is defined as an Archive, and an Archivist is someone who works in an archive. So, we're going to get rid of that in part of our tidy-up exercise. As well as all the specific policy changes, some of the changes to the legislation will deal with the fact that we've got rid of this, and some of the provisions which are hidden underneath the Act at the moment will be there for everyone to see.

It's a huge job which we enjoy considerably and are slowly getting through at the moment. So, what are the measures that we're amending? First of all, [Section 42](#) is one that already exists, it's an exception for archiving and replacing copies. We want to update it to become what we think it should be, which is essentially a preservation

exception, so that a library or archive can make copies of material in order to preserve it. At the moment this doesn't relate to all types of material, it's just literary, dramatic and musical works, according to my slide. We're going to apply it to all types of work, sound recordings and what have you, and we're also going to extend it to apply to other organisations like museums and galleries. We want to change some of the language so that it fits with the modern process of preservation. However, the exception will remain limited to situations where it's not practicable to purchase a replacement copy, and Tim might have something to say about that.

Another to do with archives are [these interesting provisions](#) which let you record a folk song or a broadcast, and we're more or less keeping those as they are, but they have a very strange structure. At the moment, you can do this if you're a not-for-profit archive, but you have to write to the Secretary of State and ask him to be added to a list. The Secretary of State has to do this is by drawing up a statutory instrument, which is legislation that needs to go through Parliament, and it takes a long time to do. Everybody on both sides of that equation finds it quite frustrating, so we're going to get rid of that process.

We are updating the provisions that allow people to make copies for the purpose of research and private study. The first one of these is something which we can all use, it allows an individual to copy material for the purpose of non-commercial research and private study, as long as it's to an extent that's fair. There's an equivalent type of provision which applies to libraries and archives, we're updating [Section 37](#), and again, this is in line with our general theme of making things work with different types of media and technology. This is going to be amended to cover all types of published work, whatever that material is, whether it is a film or a sound recording or artistic work, it'll be covered. You'll only be able to take a reasonable amount, which is currently how the Act phrases that.

We're also going to try to simplify some of the provisions around declarations that people need to make, where they promise that they won't do any bad things with the material they get hold of. Again, the general theme continues with the other provisions. We already have

many of these, but we're going to update them so they cover all types of material.

There's a new one here, which is rather unusual and describes a slightly strange concept, which is making works available on dedicated terminals on the premises of the institution, and the language for this one comes from the EU Copyright Directive. We're implementing it more or less as it is in that Directive, so it's a bit of an experiment and we will see how that works in practice. All of these exceptions are always construed narrowly and are designed to be relatively narrow, so they don't cause harm to creators. Again, this has quite a strict list of conditions on use.

There are other exceptions which we are amending or introducing which aren't specifically listed as being for libraries, archives or similar institutions, but are going to be available for people to use and will be of benefit to people working for those bodies. One is for [text and data mining](), if you have lawful access to some material, research journals, you can make copies of those in order to carry out data mining. The technology will have to make copies in order to do that, so we're going to permit it. It clearly won't let you have access to material that you haven't already got access to, so you won't be able to get access for free or circumvent access restrictions.

On education we're going to amend the current provisions so they apply to all types of material. Again, we're going to have a new ['fair dealing' exception]() for teaching purposes that people will be able to use in classrooms and lecture theatres. This is one of our classic examples of how the law has been left behind by changes in technology, because we already have a fair dealing exception which allows you to copy an extract using chalk and write it on a blackboard, without that quote infringing copyright. If I were to do the same thing in a presentation here then I would be at risk of infringing copyright, because it doesn't apply to anything which uses technology. We're going to turn that into a more general purpose, fair dealing exception. Fair dealing works to the extent that, if that activity would harm creators, then it wouldn't be permitted, so it's all minimal, reasonable uses.

Finally, another important one here is an [exception for quotation](). We already have an exception for criticism and review purposes, and again it's fair dealing, so you can only take a reasonable amount of a work. We're going to slightly expand this so that it takes advantage of the full scope of the European law, which allows an exception for purposes such as criticism and review, and many minor uses. You have to stretch the boundaries of criticism and review at the moment to use a quote, even if the use is extremely reasonable and small, so we're hoping that that will now be captured in this new provision.

We're doing a lot of other things as well. Maybe some of you will be interested in those, and you can ask me questions later. But those are the ones that I think are most relevant to your work, and I hope that you contribute to our review process and help it all come into place as it should. Thanks.

(MK) The next speaker is Nick Munn and again, I think you will introduce yourself.

(NM) Thank you very much for your continued attention. I know after lunch is not the best time to be speaking or listening, so thank you very much for bearing with Robin and for bearing with me as well. I'm Nick Munn, I am Deputy Director for Copyright at the Intellectual Property Office and perhaps of most interest to you will be my role in making sure that the UK rolls out orphan works schemes, both our own, the European Directive and an extended collective licencing scheme. I'll come on to the significance of those for archives in a minute if I may, for those of you for whom it's already not only obvious but have a long list of questions, which please do ask later.

But before I dig into the substance, I would very much like to say thank you on my own behalf, Robin's and the Office's for the invitation to come today. It's been a really fascinating event to be present at. I've certainly learnt some useful things that I've squirrelled away, some of which I may repeat now and a few that may surface later. I'd particularly like to endorse Martin's urging of hearing directly from people with an interest in things, and not just from representatives of a particular interest. I'm particularly grateful to Lesley for what I think

has to be the most polite lobbying I have seen in 4 ½ years at the Intellectual Property Office, for asking in a wonderful Scottish accent as well, "Can we please change copyright?" Well, the good news, as Robin has been telling us, is that we will be changing copyright in some respects, and I'll talk about some of the other copyright changes that we're going to be making.

But I thought I'd better start with the elephant in the room, which is the other statement of Lesley's which Ronan opened with, about copyright making the job of archives impossible. Well, it clearly doesn't make your job impossible, which is good news. You've managed to do it and there's a whole bunch of you here today which is brilliant, but it may not make it any easier and it may make it a lot harder. I think we are all agreed that that is not a desirable situation. Responsible archivists, acting in the public interest and without harm to other people's interests, are restrained from doing that in undesirable ways, or have to jump through too many hoops that are too big. Robin's explained, in the context of copyright exceptions, about taking some of those activities out of the realm of requiring explicit permission in the interests of making them happen more freely. Now that's not something we do lightly because copyright's there for a reason, it's there to protect the interests of creators of things, and it's not necessarily just there to protect the interests of creators who are deliberately creating something at the time from which they'll derive value. But that is a lot of the original intention. We are caught between the need to protect copyright as a thing which protects people who do creative things, of which the ultimate beneficiaries in many cases are archives and their users, but also to make sure that sensible things done by sensible people, and the Codebreakers Project and others that we've heard about today are very much in that category, can still happen with the minimum of fuss.

I hear very much the statement that preservation without access is pointless. It's a start. You can't access it until you've preserved it, but the end goal, not only for you as archivists but also for the government, is that people can access knowledge. I'm speaking now not just with a copyright hat on, but this is what the government and indeed

successive governments have been all about. Stuff that you cannot access, you cannot use productively; whether that's productively in the sense of things that are for society, good for research, good for the economy, or whatever else.

Ronan gave us some very interesting questions at the start of this morning, which I think are the right questions. I'm not sure I can necessarily answer them for you, but I'm going to give at least one of them back to you, because the first thing he said was how can we help archivists do their job and I think that's the right question, and I'm going to try and answer that. I'm working from the slight disadvantage that whereas Robin, I'm sure, is the person in the room who knows the most about what the government is doing on copyright exceptions, I could very well be the person in this room who knows least about archives. So, you have me at an advantage, but I'm going to do my best to try and answer this from your point of view, and not just from ours.

The other thing I want to highlight from what Ronan said was his appeal for facts about copyright budgets, how often fees get paid and to whom and for what reasons. This is really important information, without which we are going to struggle to do the right thing on a number of points. I particularly want to make that play in the context of what we're doing and planning on orphan works. Now as you'll be aware, there is an [Orphan Works Directive](#) that the European Union has approved and is going to have to come into force by the 29th of October next year. That will be another copyright exception, which for certain types of works for cultural uses - which on the whole is what archives are going to be doing, though not exclusively as we heard from the Churchill Archive - will allow digitisation on the basis of a diligent search. We had a great discussion on diligent search this morning and I've taken some very useful points away from that. In particular, one that struck home with me is the point about the diligent search depending on having knowledge of the subject, of the field, being a person who knows what diligence looks like. That's actually a really important point that we must make sure we don't lose, when we come to talk about it.

I said earlier that the government was somewhat torn in terms of protecting the rights of copyright owners, and making sure that sensible things done by archives and other people can happen. You will see that reflected in what we are doing and will continue to do on orphan works, where we are trying to balance the wish of some rights holders, particularly photographers but across a range of sectors, to make sure that their work isn't exploited unfairly, against creating a framework to allow archives and others who are responsible and, frankly, are going to do nobody any harm. I absolutely accept that's the intention of everyone in this room, to go about that with the minimum of trouble. I hope soon the government's going to be able to consult on precise rules for how we're going to implement the Orphan Works Directive and how we're going to implement the UK's own Orphan Works Scheme, which will allow for some commercial uses, and for some other types of work such as photographs to be used non-commercially as well. In the spirit of learning from people, we are going to be looking to develop some guidance around that use of orphan works in the UK, which draws on expertise in different subjects and of different types of work. Not to come up with a prescriptive set of rules, because one of the things I heard, particularly from Christie this morning, was this idea of there being no fixed rules. You couldn't just write a formula that says, this is how you do diligence, this is how you find people. You have to know what you're looking for and then go where the information takes you, and at some point you're going to have make a judgement; this isn't happening, it's just not reasonable to try and look any further. The ends are all dead ends. All of that made perfect sense.

We're looking very much to root what diligent search looks like not so much in rules, but in the judgement of people who know what they're doing. Having said which, there is a real risk that that will not provide anyone with the certainty they're looking for. Again, there's a trade off between providing people with certainty, definite-ness, and for people who are a little less expert than you in this room, some help on knowing where to start, without just prescribing something which in some cases will require you to do things that will never turn up an

author, and in other cases will be the wrong things to do, but you'd find him easily in another way, so giving wrong permissions in either case.

So, to respond again to Lesley's call, can we please change copyright? Yes, we are changing copyright. Now, what we've heard from a variety of people is, "Yes, but A, you haven't done it yet and B, you're not changing it enough or in the right ways." Okay. Well, because we haven't done it yet, it's very hard to tell how far what we are doing will go. I have a feeling that we will probably not get to the future of not having to worry too much about copyright for archives in one go. What I am hoping, though, is that between the copyright exceptions changes, the Orphan Works Scheme and the possibility - although I think it's not an immediate possibility - of some form of collective licencing around archive uses, which would require a lot of negotiation, but is starting to be put on the table by what the government is doing and providing for, I'm hoping that's going to start to move the balance towards archives being able to do more easily more things that are in their mission to do, without moving too many owners of copyright into the position of being in any way deprived or upset or harmed in their reputations – we heard about reputation earlier – in what happens.

I'd like to conclude just by saying we had a provocation from Martin about the current state of the law, and the fact that the financial sanctions, in his view, on infringing copyright for an archive project were just not that big. It's funny, I think Martin thinks he's telling me, "You know what? You shouldn't bother with all this." There is part of me however, that's writing down, should we make a criminal offence? I've had to have this conversation with the team who are working on orphan works, and I have to say, the conversation I had with them was, "I really do not want there to be a new criminal offence for so many reasons, which starts to criminalise the activity," – that's not the way this is meant to be going. Not because we are, in any sense, wanting to condone criminal behaviour. In fact a lot of what we are doing is because the government, for reasons you will understand, just can't put itself in the place where it is going to wink at, or condone breaking the law, even if it's not breaking the criminal law. If you think about that for a second, you will realise that however inconvenient and annoying and

difficult that is for archives, you do not want, as a general principle, to have a government that's got a bit of a, "Oh well," attitude to the law. You do not want to have that. And if you've ever visited a country where that is the attitude of government, you will know about it and you may indeed have stories to tell about it.

We're in the position of trying to provide legal answers. We are not able and not necessarily rushing, however, to close down the risk-managed approach. We know why you take it, we know that you're interested in doing that for your own laudable objectives, and that will continue to be an option as far as I can see, into the future. Certainly, it's not one that we're looking to close down. I would say that by creating legal alternatives, you may find that the courts are a little bit less sympathetic about not using them where they are appropriate, but equally, those options will remain open to you, and while you will understand that I cannot in any way urge you or even recommend you to use them, nor can I tell you that they're going to vanish like fairy gold.

So your position, I hope, as a result of what the government is doing, is better. You have more options. You have more options to do things legally. You have no fewer options to do things unlawfully, if you feel you must. And on that note, let me just return to the conditions of the legal scheme. People have been very keen to make sure that what the government does is as simple and easy to use as possible, so they can stay within the law and stay within what the government does, as opposed to breaking the law. What I'd like to say is that the evidence that's being collected, for instance around the Codebreakers project, is the kind of thing that will really help set appropriate conditions for licensing of orphan works through the UK Scheme. Keep collecting that, keep giving us information from your own experiences as archivists, and particularly from these projects where you've got some systematic data. That will help us condition what we are doing and get the answers, if not completely right from your point of view, at least a lot closer to right than we would have done otherwise. Today's been wonderful for that, I'd like to thank CREATe and Wellcome in particular for letting us be here. I'm going to hand over to Tim Padfield for some

thoughts on my thoughts, and things said so far, and then Martin will attempt to give us a severe grilling and let you get a few questions in as well. Thanks very much.

(TP) Hello everyone. I don't represent anyone any more. I'm in the sad position of being retired, but I'm down here as representing the International Council on Archives, which I do every six months or so. I'm here to talk about the impact of copyright on archives, and to respond to what you've heard from the IPO and some of what you haven't heard from the IPO, actually.

I think archivists are scared by copyright, which is sad but not surprising. It's largely because it's unfamiliar to them. If you go on an archives course at one of the universities, you're likely to have a talk from me on copyright which will take 2 hours, and that's the only information you'll get about copyright at all in your professional training. Also, archivists are scared because in my view, in many respects, Copyright's rather absurd. My biggest feeling of absurdity about copyright is duration, so I'm going to say something about duration even though Robin and Nick haven't said anything about it all.

We've been told, and we keep being told, that the purpose of copyright is to encourage innovation, to encourage creativity, and yet we have a duration of copyright the standard of which is 70 years from the death of the creator. Why you are giving benefits to the grandchildren and the great grandchildren of the creator, in order to encourage innovation, I really don't understand. It makes the 2039 date for the termination of copyright in unpublished literary works and some other works even more absurd, which means that 15th century works are protected by copyright, even though they weren't when copyright was created in 1709. I find it really bizarre.

However, I hope you are aware that the [Enterprise and Regulatory Reform Bill](#) has given the government power to remove at least some of the 2039 terms, and when that happens, on the assumption that it does happen in a year or so, huge volumes of copyright material in archives will cease to be protected by copyright altogether. Most 19th century

material and all pre 19th century material will cease to be protected by copyright, and I think that will be a wonderful thing.

We don't know yet how it's going to happen, or precisely when it's going to happen but I am assured that it will happen, so it's something to look forward to and keep your eyes open for. It will definitely remove at least part of the absurdity of the duration of copyright.

I don't want to spend too long, because we need to give you time to ask questions so I'll just say a few things about the other changes to copyright. Preservation copying, and indeed copying by librarians and archivists for users, applying to all types of work, is a huge improvement. Preservation copying of things like artistic works, sound recordings and films has almost certainly been happening anyway, in archives across the country. I'd be amazed if it hasn't. It'll just be nice that it will, in future, be legitimate.

You need to be aware, though, that the purpose of the preservation copying exception is preservation, the purpose is not to make the material available online. Even though you're making a digital copy, the purpose of the preservation copy is preservation. You need another exception to make that digital copy available remotely, or you need permission.

Copying of unpublished works in archives for users, once again applying to all types of work will be hugely beneficial, in fact a vital change. Once again, I'm absolutely sure that archives all across the country have been copying photographs for example, for users, and maps and so on, quite regardless of the fact that they're not allowed to. I think even some of them have been getting declaration forms for these types of copying. But once again, it'll be jolly nice if it can be done legitimately.

It's worth understanding the implications of the removal of the regulations, the wonderfully named Copyright (Librarians and Archivists) (Copying of Popular Material) Regulations 1989, with two sets of brackets, I think it's a splendid name. Anyway, they're going to be abolished, but that means that the declaration form that we all know

and love is also going to be abolished. There will no longer be a statutory declaration form. The declaration will simply be something that is in writing, so you can receive a declaration in a letter, or an email from someone. You don't have to receive a particular form. That will make life a great deal easier.

We've been told about the copying of published works. At the moment, you might not be aware, that there is no exception which allows you to copy the manuscript of a published work. If you have in your archive a manuscript of a work which has been published, it's a published work itself. The trouble is, it's not covered by either of the copying exceptions, because the librarians' exception applies to published editions of works, and the archivists' exception applies to unpublished works. There is no exception which allows the copying of a manuscript of a published work. There will be in the future, I think it's terribly important and I hope the IPO has taken this on board already, because I've told them before. It's terribly important that the exception for published works applies to archivists as well as librarians, because archivists have responsibility for a great many published works. There are lots of maps, lots of photographs and lots of manuscripts of published works in archives, so it's terribly important that you have access to this exception.

You've heard about the new exception for online access to material on the premises. The current draft of the regulations published for consultation is actually a little bit strange, because Robin said that it's based on the Information Society Directive and yet it interpolates or adds in a little bit, which is that the use must be for non-commercial research or private study, which is not in the Directive. I personally say that archivists would find it extremely difficult to police any use of an access system on their premises which was limited to non-commercial research or private study. Fair enough, supplying copies for those purposes, but controlling access simply to read material would be virtually impossible.

Then we were told a little bit about folk songs and funny things like that. This is another one of those absurdities, in my view, about copyright, that you can have a copyright in a folk song, when you have

no idea who created the work or how long ago it was created. It's a bit difficult to say that there is a rights owner in a folk song, and folk material has been the source of huge creativity. If you think of someone like Vaughn Williams or the Brothers Grimm, they produced new works based on folk material and if they couldn't have done that then we wouldn't have a lot of wonderful modern works.

I shall say just a few words about orphan works. We have two wonderful new schemes for orphan works, one from the European Union and one from the UK government. Frankly, neither of them will be a great deal of use to archivists. They will be extremely valuable to individual researchers who want to get permission for, or have licence or an exception, to use a relatively small number of works. But if you're talking about 1000 works, 10,000 works or 100,000 works, the regulations will require you to do a diligent search for the rights owner of every single one. I don't see how the orphan works exceptions will be very useful to archivists. I personally expect risk management to become more and more important, which makes this conference all the more useful. I'm looking forward to hearing, again, Peter's talk later about management of risk, because I've heard him talk about it before.

And finally, one or two words about the international scheme since I'm here as representing the IPO. I represent the IPO in Geneva, meeting with the World Intellectual Property Organisation, which is contemplating a treaty, an international treaty on exceptions to copyright in favour of libraries and archives. It's not impossible that we'll get a treaty, and it would be extraordinary and wonderful if we did. The purpose of such a thing, primarily, is to improve international dealings in copyright material so that, for example, an exception in one country applies in another country, so you can transfer a copy to another country. But it'll be a long time before we get that treaty. Thank you.

(MK) We're almost at two o'clock, but we've got quite a bit of slack in the day, and the wonderful coffee breaks and the plenary at the end, which may allow us to take up some of those issues. From my perspective, the debate shapes up. I think there's less controversy around Robin Stout's portfolio of making lawful, activities which are

happening anyway, and removing some of the garbage. It's unlikely that there are many objections to improve in this step-by-step manner, even though there are some specific issues, and I think they legitimately could come up in this context as well. But if we look at what the purpose of the day is, and assess how archives can make their material available online, then I think the orphans debate is the more dramatic intervention, and it looks like in the future we'll have essentially 3 routes. We'll have the route through the European Directive, which is an exception and then you have to comply with the diligent search of these countless databases, in the country where the work was published, or for audio visuals in the land where the headquarters of the firm is. There's a huge list there. If you want to avail yourself of that exception, you have to go through that process and the resources within the archive sector are probably not easily available. The second route, if you want to go for images, or for commercial use in the future, you will have the option to apply for a licence through the IPO or an authority which will be created; we don't know how that scheme will work. Possibly, there's a light at the end of the tunnel. If it's just a removal of liability from archives, it's possible to have just a peppercorn licence fee and the diligent search requirements would be more flexible, then it may well be a slight improvement of the situation. We still have the third route, which basically says, "What are the damages? What's the harm?" There, the risk is that courts may take a dimmer view than they have in the past, because these other schemes are available. That's the situation and the question is, is there an improvement here or not? I think the IPO would have to think extremely carefully how they structure the diligent search requirement, and also whether they can create a removal of liability scheme against a peppercorn payment. That, I think, is a hope. That's the only hope I see in this.

Okay, questions.

(AV) Hi, I have a question for the IPO. Anna Vernon from the [British Library](). My reading of the Orphan Works Directive is that it includes embedded works, so how much scope do you have, given that you've got to implement this Directive, to change that and make it less onerous

for archives and libraries to implement, and do a diligent search for each individual item?

(NM) That's an easy one to answer, though you won't like the answer, which is essentially none, whatsoever. European Directives are fought out rather extensively and what they say we have to do, so to the extent the requirement is in the Directive, and I'm afraid the requirement to search for embedded works is very firmly in the Directive, we have no choice other than to do that. At this stage, I don't think there's enough trust between rights holders and the potential users of the Orphan Works Scheme, whether the Directive or the UK's, for there to be mutual comfort around any solutions that don't involve clearing individual rights. That might change in the future, when people have had a chance to get used to these new schemes. They're not the first in the world, but they are relatively new. Sorry, Anna.

(ER) Ellie Robinson from [London School of Economics](#). I have a question about the Section 42 amendment for preservation copying. You mentioned that it would be more suited to modern preservation techniques, and I just wondered if you could clarify that a bit more?

(RS) Okay, I'll try to. I think some of this, as I said, Tim may disagree that it is suited to modern preservation techniques. There are questions around how we draft it, and we've put our legislation out to ask people whether it achieves what we're trying to do. But for instance, one of the issues is that it's not entirely clear from the language in the current provision, whether or not you can take multiple copies of a work. It seems on one reading you may only be able to take a single copy. Also, it's drafted around replacing copies and doesn't use the language of preservation, and it's something that we're working on, people have given us feedback and we're just trying to make sure that the language in the legislation matches what goes on. If you read what's currently there, it describes quite a restrictive process which people tell us isn't really what they do. Does that make sense? More widely, it'll apply to different materials, so it'll apply to films and photos and broadcasts and all these things, so from that sense as well, it's better tailored to modern media.

(PG) Paul Gibbons, [School of Oriental and African Studies](#). It's a question for Robin. I just wondered about the education exception and the definition of non-commercial fair dealing. Universities increasingly are commercial, so would lectures count as non-commercial in that context or not?

(RS) This is the type of question where it's dangerous for me to try to give a definitive answer. It's fair dealing, so one part of the equation is a judgement on the fairness of the activity, and we often give the use of materials in presentations as one of our examples, the use of an interactive whiteboard in a classroom. We often use that one because it seems pretty much like fair dealing to us, although even on that you can never quite say until it's tested. But I think that on a non-commercial side, it's the purpose of the activity which needs to be non commercial, so it's not with reference to the institution. In fact, the fair dealing provision doesn't apply to particular institutions at all, so it's all about whether you are teaching, or we use the word instruction for some reason in our legislation, but essentially it has to be teaching for a non-commercial purpose, and to the extent that it's fair to do that. It's not about the type of body. Whether or not the purpose is non-commercial, I think is a separate judgement and something that we can't –

(PG) That's more what I was wondering, does it count as non-commercial because it's what the individual is doing, but if the individual is a lecturer working for an organisation, does that make it commercial or not? Or can't you comment?

(RS) They're similar questions, aren't they, in relation to the research fair dealing which Nick's involved with, and the term non-commercial comes from the copyright directive, so in respect of this one we have to have it. The Directive says you can have this type of exception, but you can't go further than this boundary and one of the restrictions on this type is it must be non-commercial. As a legislator in an EU member state, it's very dangerous to start writing into your law, "Non-commercial means this," because in a few years' time there will be a case before the European Court where the European Court says, "Non-commercial means this." Very often where there are terms like this, we

simply go with how they're written down and people just have to place the normal meaning on those words to the extent that that is possible.

(NM) If I can just add ever so slightly to that answer. The Copyright Directive from which Robin is rightly answering, does contain a recital which is basically a preamble to the law that says what it's meant to do. That does suggest - I think it's number 42 but I always get that wrong - that you shouldn't be going too deep into wondering about fundamentals about does someone ultimately make a financial benefit from this, in terms of defining non-commercial. It is the activity itself, as Robin said, a non-commercial aim or a commercial aim. What you might do, for instance, in your academic institution, is photocopy the interesting bits of your Charter that say what your institution is there for if anyone starts asking questions about the commercial-ness or otherwise of your activity. But ultimately Robin is being cagey, not because he wishes to be unhelpful, but because it's a question of fact which ultimately would be up to the courts. Although that's a horrible answer to have to give, I'm afraid the whole of copyright ultimately comes down to the interpretation of something which, as Robin says, is a trifle complicated, both to read and also to apply.

(MK) Tim, would you like to comment on any of the issues that came up, just because the questions came the IPO's way? Is there still a question in the room? There's still two.

(CW) Cathy Williams, National Archives. Just to pick up, Robin you joked about the definition of archives, but how are you going to define the collecting institutions to whom these exceptions etc apply? Especially, since through the current legislation and what you're proposing, you scatter terms like library, archive, museum, creator, and associate them with different exceptions, where actually, aren't we talking about the nature of a collection, how it's been formed and how it's going to be used, rather than the institution?

(RS) In terms of the types of institution that we're referring to, again the Copyright Directive sets some limits. It says that libraries, archives, museums and educational establishments are the type that we can apply certain exceptions to, and in this area we're trying to, in some

ways, use the maximum scope that those provisions provide for us. That's why we're extending the types of organisation it applies to. In terms of the organisation, an archive, as I said, is already defined as an archive in the law and you just have to apply the normal meaning of that word. In many areas, part of the problem with the current copyright legislation is that it bends over backwards to try to define every last eventuality, and minutely define every last definition, and sometimes that ends up not being helpful. People constantly say, "Can you define this further, can you define that further?" The law ought to just make sense. You give words their normal meaning, the courts will give it their normal meaning, and hopefully we don't have to try to tell people what these things are.

(NM) I think that's right. Your choices here are, you can either try to persuade someone that you're really an archive, or you can have parliament decide whether you're an archive or not. I think you have a better chance of defending your own existence as an archive, and the best way of doing that is to leave 'archive' to have its natural meaning, and to let you justify your activities as well. "We're an archive, we do archive-y things, this is what an archive is, this is what we do, look, we fit." As opposed to leaving it up to people like Robin and me, to try and write down from our understanding of archives, yes we'll talk to you about it, a formulation which may not in 20 years reflect what archives are like. We think you're better off with this approach, frankly, but if you disagree you'd better tell us.

(PT) My name's Pam Temple. I'm from [UCL](UCL), but I also represent a community archive. My question is quite simple. Can you say when the flexibility around 2039 restrictions on unpublished works is going to be lifted; can you give us a timeframe for that?

(NM) What do you think? The honest answer is, no, I can't give you a date right now but I don't see any particular reason why we would hang about doing it, so watch this space. We're not thinking this is going to be a 5 year project. I would hope, but don't hold me to this please, next 12 months. If you leave me a card or some details, I'll get back to you with any sort of update.

(PT) Thank you.

(SC) Susan Corrigall, [National Records of Scotland](). I have a quick question about the Orphan Works Licencing Scheme. Do you have any idea yet of the composition of the panel who will define the bounds of a diligent search?

(NM) One, I'm not sure we would use a panel as such. Two, partly as a result of the discussion we had this morning, I'm not sure that defining the bounds is quite how we're going to want to do that. The sense that there is no one rigid rule for, "It is a diligent search if you do it this way, and it is not a diligent search if you do it that way," would mean that we want to leave some of those boundaries to sense and judgement, and to pick up a word that Robin used several times, fairness. The basic rules for the Orphan Works Scheme will be consulted on in the next little while, so I'd hope, but again don't hold me to this, it just depends on what ministers really want to do, I might hope that you'd have a chance to see that before Christmas at the very latest. So firstly, everyone gets a voice, and it's not just going to be some panel that's making this up. Secondly, if there's anyone who thinks they have a burning amount to contribute on what diligent search looks like, or doesn't look like, and we're quite keen not to over-engineer as well as make sure it's adequate, so what's a minimum standard of diligence, rather than what is everything you could possibly think of to do, then I'm happy to take people's details as someone who thinks they might have something to add. We might be coming round to ask, "How does it work for you, what does it look like, who knows about this, who can we ask?" We are very keen to get that working right when the scheme is introduced, and not in the light of bitter experience of no one using it, because they don't think it'll work.

(MK) Will you be here for the final plenary, will people have a chance to catch you again?

(NM) I'm afraid I'm going to have to shoot off, but I will certainly be here for the coffee break, or what remains of it. I'm very happy to take immediate questions now, otherwise you can contact us, usual

government email addresses, via panel members or Googling will probably work.

(MK) It's exemplary that the IPO is willing to engage with us in this way. They're not always easy conversations, but as you earlier said, generally we are quite friendly, apart from myself. Ronan, shall we close?

(RD) Yes, I think so. There's two different timelines. You've got one timeline on the back of your card and a different timeline up here and we've exactly split the difference between the two. If we wind up here, and restart at quarter to three for the final panel, and then we'll go on to the plenary.

(MK) Thank you.

PANEL 4
ARCHIVES AND COPYRIGHT:
THE VIEW FROM THE US

Speakers
Peter Jaszi (Professor of Law, American University) – hereinafter **(PJ)**
Matthew Sag (Professor of Law, Loyola University) – hereinafter **(MS)**
Peter B Hirtle (Research Fellow, Harvard University) – hereinafter **(PH)**
Chair: **Ronan Deazley** (Professor of Copyright Law, University of Glasgow) – hereinafter **(RD)**

Questions and Answers
Cathy Williams (Head of Collections Knowledge, TNA) – hereinafter **(CW)**
Tim Padfield (International Council on Archives) – hereinafter **(TP)**
Martin Kretschmer (Director of CREATe, University of Glasgow) – hereinafter **(MK)**
Lesley Richmond (University Archivist and Deputy Director, University of Glasgow Library) – hereinafter **(LR)**

(RD) It's really lovely to see so many people have stayed to the end of the day, although there are drinks still to come so I hear.

We've heard already from Tim and others, that the issues we are talking about today are not issues with a national dimension, they have a truly international dimension, and when we were planning this event I thought it was very important to hear about what is going on in other similar jurisdictions. We have three really very distinguished speakers from the States, Matt Sag, Peter Hirtle and Peter Jaszi who have come to give us a flavour of what the nature of the conversation is in the States, what is happening over there, and perhaps what we might learn from what's happening over there. I can only extend a very warm thanks to

all three, who have travelled a very long distance to be here, and one of whom has interrupted their training for the Chicago marathon.

I'll let you work out for yourselves who that is [laughter]. Peter is going to speak first, all I would say about Peter Jaszi is, for me, he's one of the finest and foremost thinkers about copyright and American copyright law of the 20th and 21st century. He is also a dear personal friend so Peter, thank you very much for being here.

(PJ) Thank you Ronan. I have practised and taught copyright for 45 years or so, and for the most part I am, as you might expect, very well disposed toward this general body of law. Copyright has been good to me and I think, by and large, copyright has been good to culture, and to society, but there is an old adage that teaches that it is always possible to have too much of a good thing. We may, I think, today be in such a position, where copyright and in particular where the demands of copyright compliance on various sectors of cultural practice are a concern. The situation is particularly difficult for a field like archival practice, which gets almost incidentally sideswiped by the general culture of copyright compliance. Today, I want to talk briefly about three topics that relate to the US situation where the problem set that you heard described so well this morning is concerned. Our situation is a little different from the situation in the UK because we are working through this set of issues, through risk analysis, that is being conducted in the shadow of the fair use doctrine. I'll talk about fair use in general, and where it stands in the United States today as I understand it. I will talk about a few examples of relevant, very recent case decisions from US courts relating to fair use which have significance for the archival enterprise, or for the performance of archival functions. Then I want to take a very quick look with you at the so-called best practices project with which I have been involved with many wonderful colleagues for the last decade or so, and to talk about how some of that work may have a bearing on these issues. But first, if I may, I want to lead off with an old copyright lawyer's reflection on the relationship between doctrine on the one hand, and risk management on the other. I agree, of course, with the point that was emphasised this morning in Christy Henshaw's slide and that a number of other speakers have confirmed

since; that decisions about copyright clearance, when to do it, how to do it and how much to do it, are always considerations based in the end on a vision of risk, and of risk tolerance in a particular institution. We can't dispute that proposition, but for myself I subscribe to the notion that risk analysis ought to proceed on the basis of as thorough as possible an understanding of the applicable law. Risk analysis conducted in the absence of such an understanding is dangerous when, without a solid understanding of the law, of what it is and of the direction in which it's moving, one runs a significant chance that in risk analysis one may either overshoot or undershoot the mark. In other words, I think that the work of trying to clarify the law and its application through interpretation, and also sometimes through advocating for better rules, is an essential input into the risk management activity, and with that let me talk about fair use.

First of all, in one sense fair use has been an integral part of the US copyright system since the middle of the 19th century, when the judges first thought it up. It's been essential because, from its origins, it's always been clear that it had a close relationship in theory to the fulfilment of the core constitutional purpose of copyright in the United States, to promote cultural flourishing. It's also been recognised that fair use is present in our law for another rather essential theoretical reason, and that is that it helps to promote and safe-guard free speech values which might otherwise be at risk through over-enforcement of copyright norms. That's all very good in theory, it's been running in the background of American law for a long time, but its emergence as a practically significant feature of our law really dates back only about 20 years, to the judicial embrace of the proposition that the cultural, expressive or communicative purpose of a significant or substantial unlicensed use of copyrighted material, is the single most significant consideration in determining whether that use should nevertheless be considered non-infringing.

So, educational uses receive significant deference and, of course, so do so-called transformative uses. This is the rubric that has emerged in our court decisions over the last twenty years or so, that has been embraced by judges in every region and at every level; the notion that

transformative use is those which involve re-using material to convey a new message, to address a new audience or to communicate substantial new informational value, are in general privileged uses. They may be educational, they may not be educational, they may be non-profit or they may be for-profit. They may be parody, or critical commentary, but they need not be. They may also be uses in the nature of illustration. So, in particular, these case decisions of the last 20 years, the accumulation of precedents, has taught us that re-contextualising copyrighted material, without modifying it, without painting moustaches on it, or shrinking it, or filtering it in some way, also can and often does qualify as transformative fair use.

Furthermore, this fair use doctrine, which is codified in [Section 107 of the 1976 Copyright Act](#), is aggressively open ended and has occasionally been criticised for being vague and unpredictable. Because it is open-ended, there's no obvious limit on how, or to what activities, present or future, it may be applied. The other thing that's important to recognise about fair use is its relationship to the other more specific exceptions that are scattered throughout the US copyright statute and, in particular, since it is of immediate concern here today, to the archival and library exceptions that are found in [Section 108](#) and which, as Peter Hirtle will explain to you in a little while, haven't been updated for a very long time and don't show any clear sign of being updated any time soon. We know, because the statute tells us so, that fair use supplements and complements these specific exceptions in Section 108, and the fact that a particular activity in which an archive may engage in the course of its digitisation programme isn't literally covered by Section 108, merely gives rise to the next question and that is whether it may in fact be covered by fair use. So, in fact, fair use can and does reach activities like preservation and access, which are at the heart of the archival enterprise and in particular, at the heart of the move toward both digital preservation and the provision of remote access to archival collections, and it can do that without regard to the specific kind and character of the copyrighted material involved, or even the exact nature of the institution engaged in archival activities.

I was interested in the last session in the discussion of the definition of an archive that will be or won't be, as the case may be, part of the new UK legislation. The wonderful thing about fair use as we have it and practise it in the United States, is that we don't have to worry about that question; any entity that performs positive cultural functions, be it a great institution like this one or be it a small personal website which curates and presents examples of the culture of the past, can qualify under this important exception.

So that's my first proposition, a little bit in general about fair use, and now I want to talk a brief amount about recent case law with a focus on one case, which is the 2012 decision of Judge Harold Baer in [Authors Guild against HathiTrust](). I have to say at the outset that I'm not an entirely objective observer or reporter here, because I am a counsel to one of the parties defended, the National Federation of the Blind. But, putting my lack of objectivity aside, here is what I think I can say and I think that others will agree with, and my colleagues will keep me honest about the decision's implications for archival practice. What went on in that case you know well, it was a spur line off along the main track of the [Google Books litigation](), the Authors Guild having declined to sue the libraries that had provided Google with, give or take, twelve million books, some 60% perhaps of which are under copyright. They decided at the last minute that they were going to try to draw the libraries in as defendants in an additional litigation. My clients, the National Federation of the Blind and several blind individuals, joined as additional parties defendants early in this litigation. The case was litigated on the issue of fair use especially, although not exclusively, as fair use analysis may be informed by other policies including the important policy of providing accessible materials to print-disabled people, which is expressed in the United States in, among other things, the [Americans with Disabilities Act](). In that case, the judge declares that this whole enterprise of the HathiTrust participating, making the copies and then using them for a variety of different purposes constituted fair use. His finding, in particular, was based on the notion that this vast and unprecedented digital collection would serve a preservation function, one that in fact goes well beyond the limited carve out for preservation in Section 108, which again Peter Hirtle is going to discuss. It enables

search and especially text mining, about which Matt Sag is going to have more to say in a moment, over very large amounts of information and the digital scans that are in the HathiTrust can dramatically facilitate the delivery of accessible copies to print-disabled students, teachers, scholars and others. This list, it's fair use because it serves preservation, it promotes new modes of search and text mining, and it promotes accessibility, was the sum total of all the justifications advanced by all the defendants in the case. Judge Baer didn't miss one of them.

Now, the decision is on appeal and it's always possible, although I don't think it's likely, that the higher court, the second circuit court of appeals, may somehow trim back that rationale of decision a bit. As it stands, this application of fair use to one particular kind of digital project has larger implications for the archival enterprise. That is, the cultural justifications that are embraced in the HathiTrust opinion are now merely examples of the kinds of justifications that could be advanced successfully, in my view, for a variety of other mass digitisation projects.

Of course, we have other advantages in the United States which you don't enjoy, in thinking about copyright as it relates to mass digitisation. One of them is that our law, at least, permits one co-owner of a copyrighted work to authorise its use for most purposes, subject to a duty to account to other co-owners. In cases of joint authorship we don't require a majority, a consensus or a unanimous approval on the part of joint owners. That makes things easier. Although United States law suffers from having a rather swinging statutory damages provision which often strikes fear in the hearts of users, and I would say both institutional and individual users, there is a nice provision written into our law that says that anyone who is in an archival or library capacity and proceeds, in good faith, to engage in reproduction, distribution or dissemination of material on the basis of fair use, can be shielded from statutory damages. Now, of course, as we know in most cases of the kind that we have been talking about today, the only actual damages that would likely to be appraised against a defendant would be nominal at best. So, we've got a lot of advantages. The courts have made our life

relatively easy in the United States in their interpretation and application of fair use doctrine. Some features of our statute also work very much in favour of the kinds of activities that we've described today, and yet still institutions behave in ways that may be, at least in my view, excessively cautious.

That brings us to the last thing that I want to talk toward today and that is the way that, in the last decade or so, a variety of practice communities in the United States have been working collectively to try to get out ahead of the courts, and to blaze new trails for the successful assertion of fair use in a variety of practice contexts. It has been my great privilege, working with a wonderful team at American University in Washington DC where I am based, including my invaluable colleague, Professor Pat Aufderheide in the School of Communication, to help with a number of these [Codes of Best Practices](), beginning way back a decade ago with documentary film makers and proceeding through the most recent of the Codes to be released, which is this one: The American Research Libraries [Code of Best Practices for Fair Use in Academic and Research Libraries](). These Codes are designed to take advantage of a feature of US fair use law that was first observed by our wonderful colleague, Mike Maddison of the University of Pittsburgh, who read all the fair use cases in the US courts from the beginnings to what must have been then about 2004, and came up with some generalisations about the patterns he saw. One of those patterns was that when US courts, faced with the problem of applying fair use to some new situation or at least new insofar as the courts are concerned, they look (when it exists) to documentation of what people in the field, whatever the field may be, think is legitimate, culturally positive, mission-related activity.

What we have been trying to do with a variety of communities in the best practices work is to convene them to go through the process of developing, for themselves, consensus statements about the ways in which the unlicensed use of copyrighted material is critical to the fulfilment of their expressive, artistic or cultural missions.

The first best practises project to deal directly with archival issues came in 2009, when the [Dance Heritage Coalition](), the preeminent

organisation of dance collections in the United States, released its statement of best practises in fair use of dance related materials, which deals with a wide range of situations in which copyright constraints come into conflict with the archival mission. Talking to the people I know and who I keep up with in the field of dance archiving, that statement has made a big difference. That statement has given many institutions the material they needed, the source of courage they needed, to take or to make appropriate risk management decisions. More recently, this code I showed you a moment ago, the 2011 ARL Code addressed digitisation for preservation, format migration and as Peter Hirtle will mention in a moment, the archival harvesting of websites.

This team is hard at work on a document that will address best practices and fair use in the archival sector more generally. This began with a focus on the problem of orphan works in archives and library special collections, but the work which should come to an end sometime at the end of this year or very early next year, has changed its focus somewhat. We are finishing two weeks from now, a series of group interviews involving more than 100 archival professionals from all over the United States, north, south, east, west at every level of activity and our next job will, of course, be to distil from all of this material some common positions. As we have gone along it has become clearer and clearer to us that the problem is not an orphan works problem as such, not a problem about establishing the appropriate standards of due diligence for search, as we had somewhat imagined it might be going in, but that it is really a larger issue about how to proceed with the mass digitisation of collections that include a mixture of public domain material, material with multi-known owners, and with material whose ownership can't be readily determined.

What is emerging from this discussion, and it would be getting ahead of the game to say a great deal more about what the outcome is going to be, is that subject to some qualifications, most US archivists and special collections librarians, believe strongly that their activities do represent significant and culturally positive transformative, educational uses that have a very close connection to overall institutional mission. There

seems to be strong support, in other words, for an expansive approach to the application of fair use in this area, which we are going to try to mirror as faithfully as possible in this document, although many of the special considerations that we've reflected in the discussion of Codebreaker's clearance policy this morning, also will be reflected. I am going to look forward very much to future developments here, and I hope that you will follow future developments in the US for whatever value they may have.

It falls to me in the end to thank the organisers of the meeting, the Wellcome Trust, CREATe and everyone who has done such a splendid job of putting on this event, for inviting me and making it possible for me to take part. Thank you very much. [Applause]

(MS) I just want to echo what Peter said, and thank you all for still being here and thank you Ronan, thank you everyone else for putting this together and, of course, to CREATe and the Wellcome Trust.

What I want to talk to you about today, is really just one slice of the digitisation question, but also to frame that in context and to suggest that, for those of you who didn't realise that you should be really concerned about data mining: you should all be really concerned about data mining.

Here is an abbreviated timeline of the [Google Books](#) project litigation. It is intensely abbreviated but essentially, Google announced its plan to start digitising libraries to make them searchable in 2004, the [Authors Guild](#) class action lawsuit followed fairly shortly after that and we got distracted for a few years on the rather ambitious settlement proposal. The settlement was rejected and then in September 2011, the Authors Guild filed a separate suit against the University libraries. That is the [HathiTrust](#) law suit. The Authors Guild case proceeded really quickly and jumped ahead of the versus Google decision, so that Judge Baer gave us a positive ruling on fair use for libraries, before the other court has had a chance to make a ruling on fair use about the Google Books project at large. In fact, we have just had oral argument on that on Monday, the transcript of which is incredibly encouraging.

My broader point: there are different kinds of digitisation programme, they each raise different legal issues and bring in different stakeholders, and this is an important thing to keep in mind. Broadly speaking, I think you can categorise them in terms of preservation, data production and analysis, and then as display - actually giving access to these works through some kind of digital vehicle. In terms of display, I think you have to slice that up even further to talk about the special issue of providing access to the visually disabled, and then scholarly access, and then general access. Depending on what kind of digitisation project you are thinking of, the implications will, quite naturally, be different. I have asterisked two of these because these are the issues that we got a positive fair use ruling on in the HathiTrust case. I think Judge Baer was quite sceptical of the preservation case, although without ruling against it.

Library digitisation for the production of data is an important strategic issue. It unites commercial and non-commercial interests, it has a strong, almost undeniable normative appeal, and it raises some fairly clear orphan works problems. Most importantly, once you accept the argument, then you accept the argument for digitising the whole collection which, again, reemphasises the orphan works problem. What this means from a litigation context, is that in Google Books for example, if you agree that it is okay to digitise a library to build a search engine, but you think a three line snippet is too long, you think that it should only be a two line snippet, then there is no class. You have to go in, and you have to look at every work and see whether it was in-snippiated or not, and that is a much more difficult lawsuit for the plaintiffs to bring. Even if you don't care that much about data production without snippets, it's still really important as an issue.

I want to discuss the legal argument, which I think works in the US and should also work in every other country. My terminology may require some local adaptation, but essentially we are talking about digitising non-digital objects in order to extract data from them, and in many cases to compile new data about them. This metadata, as just a matter of textbook copyright law, is not copyrighted. Copyright protects expression. It does not protect facts and ideas. It certainly doesn't

protect facts about a work. Everyone knows this, but it is important to remind people of all the reasons why this must be true. In the US we have a very clear idea of expression distinction in our law. We also have this thing called the merger doctrine that says, when an idea can only be expressed one way; the freedom of ideas trumps the protection of expression. More importantly, metadata or data about a work, is not substantially similar to the work. Nor does it originate with the author of the work, and I like to explain this in terms of whales and dinosaurs. This is a text analysis of Moby Dick.

The word dinosaur appears zero times in Moby Dick, the word whale appears over a thousand times, and neither of those observations are substantially similar to the underlying text, you will notice. Neither of those observations can honestly be said to originate with the author of Moby Dick, Herman Melville, although but for his creation, I could not make those comments. In case you don't like graphs, I also have the same data arrayed as a word cloud. Again, this is not substantially similar, nor does it originate with the author of the text. Having established proposition number one, we get to the slightly more difficult proposition number two, which is that a copying process that only produces metadata does not infringe. In other words, the intermediate non-expressive use here is just not copying in a relevant sense, which does require a set of metaphysical understanding of copying, sort of when a tree falls in the forest no-one hears it. In the US it's much easier; we just say, well it is fair use. It is technically copying, it is technically a reproduction, but it's not infringing.

Essentially the argument is that we recognise already that parts of a work may not be copyrightable, the facts and ideas within them, we just need to port that distinction over to our understanding of acts of infringement. The fundamental justification for this approach is that copyright is ultimately about the communication of expression to someone, to the public, and so when things are copied purely for internal computational reasons, and not to convey their expression to the public, then that is something that should not be recognised as copyright infringement.

I'm going to skip the application to fair use, because none of you should really care too much about our statutory factors, and I am going to move along to legal argument number three, which is that non-expressive use does not harm copyright owners and has great social value. Here is the explanation we used in the digital humanities brief, which Judge Baer mentioned in the [Hathi Trust] judgment and Judge Chin mentioned in oral argument in the Google Books case on Monday, it made me incredibly happy because I spent a long time getting this graph to work! [Laughter] It's a contrast using Google n-gram data of the United States is, versus the United States are. If you say the United States is, then you think of the US as a single entity. Are, by contrast, you still think of it as a collection of states. This shows the occurrence of those two expressions graphed over time, and the cross-over is pretty much dead on the Gettysburg Address, this conception of the United States as a single entity. This does not infringe the copyright in any of those underlying books. You can't just look on the index for books about the United States and compile this graph, you need to scan everything. It does not reproduce any text from the underlying works and we found this to be a compelling, yet simple example. Here is a slightly more complicated, but far more brilliant example from Matthew Jockers' book on macro-analysis. Matthew is one of the other authors of the digital humanities brief. He is an English professor. He managed to show that there is an incredible spike in Irish interest in the topic of slavery during the American civil war, but that there is no corresponding spike in British interest, and a slight increase in American interest. That is a fascinating result that demands further exploration. He did this based on scans of out of copyright works, he can't do this in the current century. He was, as he said, a slave to the public domain, which is a regrettable situation.

Matthew and I, and Jason Schultz from Berkeley now at NYU, wrote this brief on behalf of the digital humanities to focus attention on this issue as a standalone issue, to make sure that it was not lost. We wanted the court to be very clear that this is not just a Google issue, it is not just a research issue, it's not just an internet issue, it is an issue with profound implications, and we gave them powerful examples that were tied directly to understanding literature, so that they would get that

this was something that was not antagonistic to copyright. Of course if you understand the way search engines work, you realise that this conclusion is vital for search engines.

I was puzzled to see that the proposed legislation in the UK will still make commercial plagiarism detection software illegal. Commercial internet search engines are presumably still illegal because they don't fall under your new proposed data mining exception. Software that searches for copyright infringement will necessarily still be illegal, which is unfortunate and ironic.

In the conclusion of the Hathi Trust case, Judge Baer acknowledged that these facts were novel, but said in a quote that just made all of us so happy, that essentially he could not imagine a version of fair use that would make this valuable contribution to knowledge, and to giving access to disabled people, illegal. I've also put up a couple of other quotes from the brief. The nice thing about writing this brief was I got to work closely with people who aren't lawyers, people from linguistics, computer science, English literature, history. We also got to write essentially the same brief three times, so that made me look very productive. We also wrote a two page comment that was published in Nature, which certainly was an added bonus and we got some positive feedback from the courts, which is always nice.

What's really nice is to be able to advocate a position that I think is so important, and to feel that we had at least a little bit of a contribution to make. I want to conclude by acknowledging that your precise legal issues are different. Unfortunately, although fair use started in England, you abandoned it about 100 years ago. The underlying policy questions I think are universal. Everyone recognises that copyright is not about facts and ideas, that it really is about expression, and I think that we all recognise the promise of big data and the problems of orphan works. I think that the challenge for this community is on conveying the importance of this to the relevant bodies, which I think here are legislators and the gentlemen from the IPO, as much as they are the courts.

In terms of action items, I really want to stress, do not be in such a hurry to leave your commercial brethren out in the cold. If you want these exceptions purely for non-commercial activity then they will always be treated as a subsidy. Subsidies don't last forever. Also, if you want there to be libraries to scan, to perform digital humanities research, someone has to scan them and it's expensive to do it en-mass, so again, don't be too quick to embrace solutions that only work for non-commercial entities.

And the other lesson here is the advantage of flexible limitations and exceptions. In the US, the HathiTrust case is just a culmination of a series of cases dealing with the internet search engines, plagiarism detection software, etc. It was not too difficult for the court. I don't think that that's the same when you have a more restrictive approach.

Finally, I just wanted to pop up a slide to say what I think the open and closed issues are now in archival and digitisation work in the US. Preservation, still an open issue, but the signs are not encouraging. Orphan works display completely open, not ripe for adjudication in the HathiTrust case. Disability access: I see that as something that has been won, I don't think that's getting overturned on appeal. Data mining: I'm pretty confident about as well. Library copies as a quid pro quo: By this I mean Google's agreement with the libraries was, we will scan your collections and in consideration we will give you access to digital files. The reason that works in the US is because we have a voluntary copy doctrine. Google says we are not actually copying, we're just making it possible for the libraries to copy, and the libraries are fair users, which completely works under US law. They could still be liable as a secondary party, but they're not, because the libraries are not infringing.

Making and retaining expressive copies: The Authors Guild has just started to argue that even if you need to scan something to make a disabled access version, you should immediately destroy the original scan once you are done. So, it's okay for blind people to be able to listen to books, it just has to be as expensive as possible, and to redo the process every single time. That honestly appears to be the message. It's still open; I am expecting that one to get shut down. Snippet display:

The most boring copyright discussion in the world because ultimately, you have to accept that there is some length of acceptable snippet, so it is just an argument about whether it is two lines or three lines. It seems completely asinine, but it's still open and we are expecting a ruling on that soon from the district court.

There are a whole bunch of really fascinating, procedural and class action questions that are of no interest to you whatsoever, so I will leave them. And then I just have an aside, some further reading. This picture is from Matthew Jocker's book, it's really interesting, if you go and read the article in Nature. Thank you very much. [Applause]

(PH) As the last formal speaker of the day, it's appropriate to return to the morning and where Christy Henshaw started us off: how can you legally digitise collections? As an archivist who used to run the Cornell Institute for Digital Collections, that is where I approach all of this. I see three ways of doing this; the first is getting permission, and this is an approach that many people, including some senior staff members of the Copyright Office in the United States have said is what archivists should be doing. I have been told that it is illegal to make a copy of any unpublished item, even for research purposes, even though we have an exception in copyright law for libraries and archivists that allows copying for users if the item is unpublished, because in the view of the Copyright Office the exclusive right of first publication is paramount. Only in very, very rare situations can someone reproduce an item for a user without the permission of the person who created it. So, the solution in that approach is to get permissions for everyone. We've had, because of some of the orphan works proposals in the US, research on what the permission process is like, and I really like Maggie Dickson's article on the papers of Senator Thomas Watson, an obscure Georgia senator who died in 1922. There's a Thomas Watson who was president of IBM, this is not him. I had never heard of Thomas Watson before, Peter, have you? They went through his papers, found that there were 3,304 unique correspondents, and they looked up those names in research services and found that 21% had died before 1937, so [they were] in the public domain. Of everyone else, they could identify four current copyright owners and wrote to them, all four. The three

responses they got back said, "Sure, go ahead, that's wonderful," and [they] didn't get a reply from the fourth. With the results, the figures, the staff time and everything else, it was over a US$1,000 a linear foot of archival material to get the result that 21% is in the public domain. You could sit down and say in that environment, that says the right of first publication is paramount, do you then spend that much money and only put up the 21% that's in the public domain, plus the three correspondents for whom you got permission? This idea of diligent search for orphan works for archives just isn't going to work, it's way, way too expensive to start off with.

If we can't use a permission model to make all of our digital collections available, we can change the law and there are exceptions, as I said, and Peter has mentioned as well, in Section 108 of the Copyright Act for Libraries and Archives. In 2005, a study group was put together to look at that section and say how should it be amended for the digital environment, and I was fortunate enough to be part of that study group for three years. The thought was that we would work very quickly and clean up some of the really odd anomalies that were in there. For example, you were allowed to make three copies for preservation purposes, that was thinking of the microfilm world where you could have a camera negative, a print master and a service copy. In the digital environment that doesn't work, we're stuck with the three copy limitation. There is a provision for digital access, but incredibly stupidly, it's limited to a terminal on the premises of a library or archives, and no-one thinks that that's a good idea, except for the EU. [Laughter]

After three years, as a group of people from cultural heritage institutions and rights owners, meeting every two months, we came up with the most innocuous, bland, inoffensive set of recommendations that we could do, which really reflect...you know, we went in and said all the time this is what we've been doing for the past ten years, and it hasn't bothered any of you yet. It is just not technically in the law right now. So we said museums should be protected, remove that three copy limit, there's a provision that says you can make a preservation copy if something is damaged, deteriorating, lost or stolen, we always love how

you make copies of things that are lost or stolen, but we said let's add fragile on to a criteria there. Create a preservation only exception that you can make a preservation copy whenever you want, if you think it's a good idea, but then there would have to be some sort of triggers or other things to actually be able to use it. Allow capture and preservation of websites, what the internet archive has been doing for over a decade, right, and we said maybe we should give some legal recommendations to this. About as innocuous as it could possibly be and, of course, that was in 2008 the final report came out, we've seen no legislation introduced as yet even though the Copyright Office keeps on saying, well it'll be this year and this year. It is at the point now where I'm a little bit worried about if it comes. The good news is that I think our report has been read and may have some impact on the shape of the WIPO Treaty that's under discussion and the categories there, but what I'm hearing in the US is we have to think about the price of getting copyright legislation. I was really struck by one of our members of our working group from the Rights Holders Community saying, well Peter, yes this is just what you've been doing, this is all terribly innocuous, but if you want copyright law to change, you are going to have to give something up. We just can't make it, so it makes sense, but there's going to be a price and a price that you have to pay and we're sitting down, and saying, oh if we open up copyright law inter-library loan is going to go away and some of the other things that we think that are very important.

So, now there is a big discussion in our community about whether it's better to live with a bad law than it is to try to change the law, because in the environment we have now changing copyright law is only going to make things worse. Isn't that a cynical view? So, what we are looking at instead is thinking about the risk management approach, and as Peter mentioned we have a real plus that there are some existing limitations in US law regarding the infringement of unpublished and unregistered works. He mentioned the fair use things, but basically you cannot get hit with any of the terrible economic damages associated with copyright, the statutory damages, but what he didn't mention was unfair use, in that exception is you can still be hit with attorney's fees and that's really what kills you, is attorneys make so much money, and

under US law if it is about something that's never been registered with the Copyright Office then you do not have to pay the other side's attorney's fees and that makes suing you really, really unattractive for any lawyer, you have to be a crazy person to do that.

In terms of legal, financial risk for archival institutions it's really very, very low. The biggest problem we have is our professional standards which say that archivists respect copyright law and they don't say archivists respect copyright law unless they think that they are unlikely to be sued and then they'll do whatever they want. We have to think about ways of respecting copyright law and still doing our mission, and one way of doing this is to rely more and more upon fair use and the ARL Code of Best Practice and Peter mentioned that Principle Four has it's a fair use to create digital versions of a library's special collections and archives and to make these versions electronically accessible in appropriate contexts. Don't just accept that, you have to go and read the limitations and the exceptions and other things with that, but it's an interesting idea that we can rely upon. There is a principle that says it's a fair use to create topically based collections of websites and to make them available for scholarly use. Now you still have to sit down and figure out are you going to respect robots.text files, what are you going to do from there, but it's in effect sitting down and saying we don't need an exception in Section 108 to allow you to do this and do what the internet archive's already been doing, but instead let's say there that, under the law as it exists now, we can go ahead and do it anyway until someone comes along and sues us.

So how does this work out in practice? Well, here's a collection we put up at Cornell. The [President Barack Obama visual iconography collection](), and when Barack Obama had his first inauguration there was a tremendous outpouring of ephemera and Trotsky's and just amazing documentation generated by the joy of having him as president and we had someone who went to Washington DC to the inauguration, and bought as much of this stuff as possible being sold by vendors on little stands, taking photographs from who knows where and putting it together and generating their own posters and printing them off in the little copy shop on the corner, and you get all sorts of

strange, wonderful things going on and I sit down and say, well there's a photograph of Barack Obama, where did that photograph come from? Did Mani Garcia take that photograph and this is likely to lead to a lawsuit for infringement? Who made that poster and sold it, it's copyrighted, but we've scanned them all, we have them up on a publicly accessible collection and what we have is a copyright statement that says the copyright status of most of these things is unknown, but we are providing access to the materials as a digital aggregate under an assertion of fair use for a non-commercial educational use. We also say in the last sentence we'd like to learn more about these items in our collections. We used to say we'd like to hear from copyright owners and follow the [Digital Millennium Copyright Act Notice and take down provisions](#) and we decided that we don't really want to have people suing us, or thinking that we're getting into a legal argument, it seems much nicer to just say we'd like to learn more about these things and if you're the copyright owner, okay! And so far no reaction to that yet.

A group from the research libraries in North Carolina has put together a kind of strategy for dealing with digitising entire manuscript collections and archival record groups and putting them on-line and trying to come up with a kind of legal justification for doing this. Then we have some people who are really radical, the [Archives of American Art](#), which collects the papers of artists, has over 100 digital collections where they have just gone in and digitised everything in the collections, in its entirety, without doing any investigation of copyright owners, without having any explanation of why they're doing it, without claiming that it's a fair use, they are just doing it all and putting it all on-line and you can pop-up, Tim has seen this before we used it up in Leeds, a letter from the Director of City Art Gallery from Leeds from 1935. This is the sort of thing that's in there and the Archives of American Art said yeah, it's really good, we'll put it up, and as far as I know they haven't got a single take down request from this yet. This was from artists who you would think would have some interest in their commercial activity. So, they push it even further.

In these situations, repositories are letting their mission rise to the fore, they are recognising that there is a theoretical danger, but it's extremely

unlikely. There is no instance that I know of, except for one very weird outlier case where an archives has been sued in the United States for ever making a copy of an archival item. Instead we're saying that archives should think about their mission. Their mission to do education, to promote the use of historical materials, to encourage new scholarship in art, and that rather than worrying about the law, you should let your professional judgement guide you on whether something is good or not.

So, I tell people, don't ask is it legal, ask who's going to be angry if I do this. Who will benefit and especially _who_ will benefit becomes very important. If no-one is going to be angry and someone is going to benefit, don't spend so much time thinking about whether it's legal or not. You can take Peter's ARL Guidelines and call it a fair use.

So, there are things you can do to minimise the potential harm, you can have the take down procedures like Christy talked about this morning, you can have a notice asking for information, if you need to, you can restrict the size of the images or do other things, and I do think it's important that we try to tell our users as much as we can about the copyright owners, and the requirements that you know about the rights and the material, and if you don't know who owns the copyright in something, then say so, don't hide it away. So, there we are, that's the take-aways, thank you.

[Applause]

(RD) Thank you very much. Three really rich, useful presentations I think, I have been making some notes that I might refer to later, but I'll open it up to the floor for questions. I think we've got about ten, fifteen minutes and then we'll have a general wrap up plenary, so I'll just open it up if there are questions, or comments or observations and if you have got something to say, please identify who you are and who you represent.

(MK) It's really a question to the archivists in the audience, how do you feel about what you've just heard? The encouragement to unlawful use, really?

(PH) It's fair use!

(MK) But we've got no fair use. So, the only thing under UK law, I don't know common law, you could look at the right of way, right to roam. There may be something, if somebody says okay you can't possibly condone unlawful behaviour, there may be some doctrine somewhere one could look to see if it was possible to try to establish a custom, but I just wonder what the reaction is by the archivists?

(PH) Okay Tim, are we nuts?

(Q1) Can someone please give more information about the HathiTrust case?

(TP) I know what the HathiTrust case is, but I think it would be helpful for people here that might not have heard about it.

(PH) I want you to answer Martin's question about whether fair dealing and the UK law would allow for a more expansive risk management approach in England.

(TP) No, the law wouldn't allow for it at all. But as I said when I was talking earlier, I think archivists are going to have to accept risk if they want to do things that they want to do, and that more to the point, the politicians want them to do. My experience is that archivists' masters expect people to be making material available and the public expects it, which means the politicians expect it.

(PH) I'm discouraged by that, because American archivists are as conservative and risk-adverse as we can be, and at least we've got the law to fall back upon and we're asking you to be even more risky without that cushion.

(MS) The lack of a general phased provision in the UK is obviously a problem, but I think perhaps a bigger problem is just a misunderstanding of copyright. Copyright is a private right. What copyright means is as a copyright owner you have the option to take enforcement action, you also have the option to do nothing. Where we see people effectively choosing to do nothing, it puzzles me that other people want to get in and defend their rights for the sake of defending

them. Copying without permission, throughout history, generally has been good. It is good, we restrict it largely in order to give artists some reward and some control. But when people don't exercise their private rights, then we shouldn't get exercised about it.

(RD) Lesley. This is the problem with me sitting up here, there is now only one person with a mic in the room.

(LR) I would agree with you because as I have already said, I have a large appetite for risk, but you still have to be able to identify those right owners who have actually exercised their rights, so you still have to do some diligence.

(MS) Absolutely, I would agree, I just think that once you've made what in your context is a reasonable effort to find the people who if asked, plausibly might have a problem with it, then you should just go ahead with the rest. And when the gentlemen from the IPO suggested that if he sees too much harmless, unobjected-to, technical copyright infringement that he might have to make it criminal, I mean I honestly thought that he was joking until I figured out that he wasn't. [Laughter]

(LR) But I think that is just where we are in the UK, archive professionals just don't really know how to do that due diligence, and that is something we will just have to deal with soon.

(RD) I think one of the things that occurs to me, it was one of the reasons why I wanted to have the input of the speakers that we have, is that you can see that they have a flexible exception that is potentially delivering quite a lot of scope and latitude and capaciousness, in terms of engaging in digitisation projects around issues of preservation and access and so on, and we don't have that in the UK and I think it's unlikely that we ever will. That means, I think, as a community of practitioners in the UK, your tolerance for risk may need to be higher because we won't, other than uprooting and moving to a better jurisdiction or outsourcing all our digitisation projects to the States, I don't want to do that, but there are some jurisdictions elsewhere in the world, that are learning the lessons of the benefits of an open-ended, flexible, fair use provision. Most currently Australia is having that

debate, there are some people in Europe that are trying to push that agenda, but I don't see it ever taking, I just can't see it and that's an issue that archivists need to think about. So I will just put that out. We've got a question at the back.

(PH) While we're waiting for the mic to get back there, I just want to talk on this due diligence searching a little bit. I think Christy hit on it and the Codebreakers project is doing it right. There is the kind of diligence you can do of trying to identify every copyright owner, as we saw in the Watson papers, and that isn't going to work. There's the kind of due diligence that's been called for in the orphan works project and the Society of American Archivists, with Peter's help tried to sit down and come up with some guidelines on what would constitute due diligence from there, and boy it was really squishy and really hard and became too much of it depends and where do you draw the line, it just becomes really difficult. Then there is the due diligence of sitting down and saying if you are in a risk assessment environment, who is likely going to be angry about what we're doing and in the Thomas Watson papers, if there's only four people you can identify, you could sit down and say, "Oh, is there a letter from William Faulkner to this Georgia Senator," or something like that, but if there's not anyone that's at all prominent, then don't bother dealing with them. As far as the orphan works approach now, the Society of American Archivists is saying due diligence is not going to work for it and collective licensing for an orphan works solution. The only hope we can see is the imposition of formalities. Formalities where you have to register in order to be able to bring a legal action, and formalities are just a wonderful idea in copyright, and I don't know how we lost them.

(RD) I feel like they're gone for good, but maybe I'm wrong about that.

(CW) It's just a random thought, the only reason that most of this material still exists is because we are looking after it, so where is there strength in the argument that it wouldn't even exist if we hadn't kept those unpublished manuscripts according to best practice archival activities, keeping them to a PD5454 standards, all of those things. Is there nothing we could do, as a sector to push back, and say, "Hang on,

it's all about our making the effort when we've been making the effort for centuries, and we've kept it safe in the first place?"

OPEN DISCUSSION WITH AUDIENCE

Respondents

Chair: **Ronan Deazley** (Professor of Copyright Law, University of Glasgow) – hereinafter **(RD)**

Susan Corrigall (Copyright Officer at the National Records of Scotland) – hereinafter **(SC)**

Cathy Williams (Head of Collections Knowledge, TNA) – hereinafter **(CW)**

Anna Vernon (Licensing and Copyright Assurance Manager, British Library – hereinafter **(AV)**

Lesley Richmond (University Archivist and Deputy Director, University of Glasgow Library) – hereinafter **(LR)**

Judith Etherton (Freelance Archivist) – hereinafter **(JE)**

Tim Padfield (International Council on Archives) – hereinafter **(TP)**

Natalie Adams (Information Services Manager, Churchill Archives Centre) – hereinafter **(NA)**

Peter Jaszi (Professor of Law, American University) – hereinafter **(PJ)**

Matthew Sag (Professor of Law, Loyola University) – hereinafter **(MS)**

Peter B Hirtle (Research Fellow, Harvard University) – hereinafter **(PH)**

Victoria Stobo (Research Assistant and Postgraduate Researcher, CREATe, University of Glasgow) – hereinafter **(VS)**

Rachel MacGregor (Senior Archivist, Library of Birmingham) – hereinafter **(RM)**

(RD) The next steps plenary has just begun. What is it that you can do as sector to begin to push back on some of these issues? Let me tell you a little bit about… the scoping project was really the first stage in a longer project that Victoria will be undertaking as you all heard earlier on. She is in week one of a PhD. We have no idea, and we haven't begun to think about project design yet, but this has been enormously useful

actually in terms of shaping, well I'm speaking for Victoria, but shaping Victoria's thoughts on that. But that's a long term project, so that's a three year endeavour in which Victoria is going to hope to engage with the archive community in the UK and elsewhere, and think about exactly these issues. But we have also heard today, Nick Munn wants to hear from archivists and too often the archival voice just isn't being heard by policymakers and by the legislature. You are absolutely right, Matt, it's not really the judges that we need to be talking to here, it's the policymakers. He's asked for input from this sector, at the moment we've got the exceptions that have gone through their technical review and it is almost certainly too late to begin to... what we are getting I think we're getting, but there are other things coming up. The orphan works deadline, implementation October 2014, how will the licencing body, the licencing authority, how will it understand what it means to talk about diligent search in relation to the digitisation of archival collections, as opposed to library collections. I think those are two very different things. But if we are not putting something on the table for them to think about, in the way that Peter has done through his best practices work, talking to stakeholder communities, getting those communities to develop and articulate their norms of practice, what is it that we think is fair use? We can't articulate those norms in the UK because we don't have fair use, but we can begin to articulate norms around what a reasonable fee is when you're trying to clear rights in an orphan work as part of an archival project. Probably zero, probably. What does it mean to engage in diligent search for archivists trying to make this material accessible for non-commercial or perhaps commercial reasons? Are there different standards that we need to think of? But that needs to come from the community; it follows on from Martin's question, it pays heed to the advice that we're getting here. Peter used a phrase that I'm going to start using and I'm going to start encouraging archivists to use and that is – "Well, technically it's not in the law right now, but we hope it will be at some future point." That's a good phrase to use, but the question is, you're here as representatives of the archive community, what are the next steps? How do we push back? Thoughts, comments, and again if we can use mics. Victoria, is there a second mic?

(AV) It's Anna Vernon again from the British Library, I wouldn't necessarily presume that the exceptions that are proposed are going to go through as they stand, and that is my biggest fear really, that we need to express our support quite vocally.

(RD) You mean they may not go through at all?

(AV) Well, they might not go through as they currently stand, because the right holders have so much access to ministers and politicians, so they might not happen in their current form really and that's my worry, that we are not articulating how important they are.

(RD) But then we have a tighter deadline than we thought, in terms of collective action. Yes?

(JE) I think in England we base a lot of our legal judgments on case law and I'm sitting here and I'm thinking as an archivist, I am aghast that the Women's Authors or whatever, put up the whole of their collection without checking anything at all. That would frighten me to death, but if we've got to start somewhere, the Wellcome has set parameters for their project and maybe the way forward for us, here in the UK, is to take what they've done and start doing a whole series of mass digitisation projects. Based on how they have done it, come together as a consensus, don't check everything, take a bit of risk, or get a little bit encouraged that the risk is okay, because at the moment I don't think we've got any case law, and then by that time we'll have a head of steam. Then if somebody does sue somebody else, we'll actually have the proof that this has been going on for a while, and we've got lots of cases and people might come down then in our favour, if you see what I mean. We just need to get started and do a lot more.

(MS) While the microphone is going back, I will just throw in a quick comment. My biggest concern is that you move from an environment where you have to make these difficult risk assessments, to an environment where you are not allowed to make risk assessments, where the only way to do this is through some complicated, bureaucratic, statutory process that has such high administrative costs that are so inapplicable to your particular archives and its needs, that it

is literally worse than nothing. There is a huge danger than any kind of extended collective licensing scheme, just becomes an extended arbitrary tax regime. The more of these projects that there are in evidence that demonstrate that the median licence fee is zero, and that these things can be done without the world falling apart, the better. Obviously then it is important that they are all done well, for that exact same reason.

(PJ) Could I add one thing and that is that what has just been said makes a great deal of sense to me, and it does also seem to me, from what I heard this morning, that the practices that are emerging in the Wellcome projects could be a very good way to start. I would have one piece of advice about how to articulate that approach and that is when explaining, for example, why you don't check the location of every copyright owner, why you make decisions about which copyright owners it makes sense to check, I would at least consider, for external consumption, casting those policies in some frame other than risk. I would urge you to consider casting those statements of policy, in terms of what archivists believe represents an appropriate level of fidelity to the underlying values of the copyright system. Because if you present to the world a set of practices, and then explain that the reason you are doing these things is because you think they are what you can get away with, that is not impressive and it could even lead to some backlash. As I tried to say before, I don't doubt for a moment that there is an important element of risk assessment here, but I think that to frame the working principles on which you proceed, entirely in that vocabulary, could well be self-defeating.

(MS) Peter would you say that they should say when they are doing diligence, that they are just confirming their implied licence?

(PJ) That's one possibility, I think. I don't have that articulation. What I would say is that what we are doing here is what we think we can do to advance the interests that are represented in the copyright law consistent with the fulfilment of our cultural mission. We are making that balance and whether the law literally gives you a warrant to make that balance, is quite irrelevant to that articulation. Again what I

wouldn't say, is that this is what we are doing because this is what we think we can do without getting into trouble.

(NA) Natalie Adams from Churchill Archive Centre again, I completely agree we should build on the lessons from the Wellcome, one of the practical things we can do, I think, is I worked on the clearance of the Churchill papers and the Wellcome worked on the Codebreakers clearance, I have got a database of copyright holders with a lot of contact details, and so have the Wellcome. We need to make it easier when we are doing our due diligence and we've decided to contact people, we need to make it easier for people to contact the people who someone has traced. Because of data protection, I can't see that I can publish my database of contact details, but we should be sharing that information, even if I am forwarding approaches for people, I think that would be a really practically useful thing.

(RD) Yeah, thank you for that.

(PH) Well of course, in theory, you should be able to give it to David Sutton and [the WATCH project](#) and they do then secure permission from the contact listed in the database, in order to be able to have that information listed there, so it is all done with permission. So, wouldn't that protect it?

(NA) I think I could explore that, I think the different sorts of people I have in my database, they are not really those writers and authors who've created a work for gain, they are more the A N Other people, so I'm not sure whether they are a natural fit with that database, but maybe that is something that we could do, take that into a broader direction?

(PH) Well it does say writers, artists and copyright holders, that last one, but I don't know whether that project of literary collections, how interested they are in expanding the scope of it, do you know Tim?

(TP) Yes, David Sutton certainly does want to expand it, and I think the answer is talk to David Sutton, ask if he's happy to take the material on and I'd be very surprised if he weren't.

(RD) Step one? Yes.

(SC) I was just going to say my experience of using the WATCH database in the National Records of Scotland is in over ten years, I've found one. And I've been looking up some people who, to me, were well know, very well known, quite productive, prodigious producers of works. So, I think the usefulness of the WATCH file - it very much depends on the nature of your collections, but that is perhaps an argument for expanding it.

(PH) Well, this just points out the need for registries and I think even in Europe with things like the Arrow Project, there's a recognition that it's really useful to know who copyright owners are, and to me the ideal orphan works system would be to sit down and say let's run our 3,000 names against a registry, and if there's a hit that's great, and if not, you can go ahead and use it for non-commercial purposes until such time as somebody wants to add their name to the registry, but again I'm dreaming there.

(RD) But again Arrow is published works.

(PH) Published works.

(RD) Yeah so there is a real problem of fit in terms of...

(PH) But not really, because why not have a registry for unpublished works as well, if you don't want to have your things be available for non-commercial scholarly use, then you add your name to the registry, but otherwise until you do, it's there. Formalities, that's a solution, right? [Laughter]

(TP) In a sense we are getting formalities and we've got the [Copyright Hub](#) which is currently in the process of being created in the UK and what it's intending to do is provide links to databases of rights owners and material in which they own rights, so there actually will be nothing to stop the archival community setting up a database of rights owners, and seeking to have it accessible through the hub.

(RM) Rachel Macgregor, [Library of Birmingham](#). Obviously, this idea of having a database of rights owners is great for certain collections, but

we are going back to the kind of ...and I was interested to see the iconography of the Barak Obama collection, we have got a lot of collections like that produced by unknown people, representing things happening in our city, different communities, so this is good to a point, but we still need to move forward on lots of modern collections which there is no point even going there with a rights registry because it is not going to fit that need at all. I don't have any suggestions as to what you are supposed to do, but just sort of bringing us back to that point, and I think it will be good to continue that discussion around how we do that. I think it is using guidelines, I am sure it is, but I think the community, the archival community does need to take the lead on that.

(RD) Can I just ask one other question then, what is the appropriate forum or fora for having these discussions and who should take the lead? Lesley?

(LR) You are going to dark waters here.

(VS) What about the [Library and Archives Copyright Alliance](#)?

(RD) Yeah, sure.

(VS) Possibly, I'm not sure. Anyone from the Library and Archives Copyright Alliance in the room?

(TP) I'm glad to say I am no longer the chairman of it. Naomi Korn has taken over as chair of LACA. And she is trying to be terribly active, but how active she can be must be limited simply as a consultant as well and has to do a certain amount of work to bring in money. But yes LACA certainly exists, and could be a forum for publishing material, I'm not sure how much actual work it could do. It might set up meetings or something of that sort, but you'd need to talk to Naomi to start with, or if people wanted me to I could take it to the next meeting which is the end of October I think, or early November.

(RD) I think that would be a very useful thing to do and maybe that is something we can talk about Tim, and the [National Archives](#), do they have a role to play in this? That's a leading question. Maybe I'll leave that hanging in the air.

(LR) There is certainly the [Archive and Records Association](#).

(RD) Sure, of course.

(LR) If he [David Mander] was still here, he would.

(RD) He would be saying yes. Yeah, I think that's absolutely right.

(PJ) So could I ask a related question? Because one of the things you heard, and I actually think that it was incredibly powerful and is incredibly powerful, and that in his modesty Matt Sag may not have explained how powerful it was, you heard the story that the impact that this brief by scholars had in these various copyright litigations. The effect that having the users of these services come forward and say this stuff is actually very important in order for us to do the research that we can do, and to produce the benefits for the culture that that research will yield, we need a certain level of archival access. We need our archives in effect to be doing certain things, and I'm very curious whether in your view it might be possible to mobilise a somewhat larger community of interests, including the various academic, commercial and other consumers of your services, around these issues. Because I think as long as it is only archivists, no matter how articulate and no matter how persuasive, this may seem to policymakers, ironically the very policymakers who are on the other hand pressuring you to make more material available on-line, like a form of special pleading. If it would be possible to broaden the base of whatever campaign you have in mind, I think it might add considerably to its effectiveness.

(RD) So this is one of the reasons why I described Peter Jaszi as one of the foremost thinkers on copyright earlier, and I often ask myself in situations like this, well what would Peter Jaszi do? He is unerring with his advice, and mentoring, and guidance, and I think that is really something that as a community we should also take on board, in terms of thinking about what our next steps are.

(SC) [National Records of Scotland](#), you mentioned the National Archives, Ronan. I'm going to make an observation in terms of policy here. I work for a government body, and that gives me and my

institution certain problems in relation to risk, because obviously the risk is minimal in terms of being sued and damages and all of that, I completely get that. However, I work for the government, I am a civil servant, we pass laws and it's a pretty tricky one for us to buy that risk, to knowingly break the law on the one hand and on the other hand I am thinking of things like Archives Accreditation which is kind of, I'm just looking to see if Kathy Williams is still here, coming from central government archives community and one of the things in there, for instance, is demonstrable respect for third party rights, and what I am seeing here, is there is a divergence. On the one hand there are arguments for increased appetite for risk for understandable reasons, but on the other hand there are these policies and frameworks and so on, coming from one arm or another arm of central government, which have to respect the law however silly or trivial the law might be. I don't have any answers to this divergence, but I think it's something that we can't forget as a community either.

(RD) I absolutely see that, but obviously as a community there are, in terms of articulating norms of practice around due diligence, that are not about appetite for risk and or, just going ahead and digitising, there is a role that the whole community clearly can come together and begin to think about and discuss those issues with a view to trying to set out some sensible policies that the policymakers might take on board, or the new licensing authority might take on board when considering what's an appropriate fee to charge archivists, whether public sector, private sector, in engaging in activities like this. You are right, there are two streams, I think, that we might begin to think about or begin to progress work on, but I mean, certainly any kind of debates around those sorts of issues that didn't include input from the National Archives, I think it is really important that you are involved in those processes. Tim?

(TP) Can I just say one thing in response to what Peter said about high profile people and so on. In Marrakesh a few months ago, we had an international treaty on exceptions to copyright in favour of the visually disabled, and that was significantly helped by the presence of a person called Stevie Wonder, who supported the campaign and he went to

Marrakesh and he performed there and so on. If any archivists can think of a Stevie Wonder in the archival world then it might help, but certainly at WIPO in Geneva we and the librarians are trying to think of people who could come to Geneva and take a fairly high profile part in negotiations, or at least in trying to persuade member states that this is something worth doing.

(RD) Thanks Tim.

(CW) Just to pick up from my colleague from the NRS, TNA is in a slightly different position because we are both an agency of the [Ministry of Justice](#) and therefore directly linked to central government, but we also have mandate from the [DCMS](#) (Department for Culture, Media and Sport) and we have done over the last two years, to be acting as leaders for the archive sector. So, we tread a really quite interesting path.   We would advocate, I would say, for something that is proportionate, hence my comments before about this stuff wouldn't exist anyway, if we weren't doing a damn good job of keeping it.   So TNA's role, yes it's a fine line, but TNA's role is to engage or support or facilitate, pick a word, conversations like this. To broker partnerships between members of the sector, between the sector and other key sectors, so I absolutely agree with bringing in support from our legal colleagues, ensuring that archives are working right across the Cultural Heritage sector, and our role really is to try and create some kind of hub for these discussions, even if we have to sort of then step aside and let the sector speak for itself, but that is TNA's dual role and we do find it quite challenging, but our access to the corridors of power, in some ways, can help us to support these conversations.   So, do use us.   At the moment we are gathering feedback through a consultation on how you all think we're doing as archive sector leaders, and I think one of the practical things you could do is all shout about, you're doing great by the way - add that first, but copyright.   So, we will respond and it gives us weight to open those conversations with government. We are supposed to tell government what the impacts of policy is having on you guys, and feeding it back to them, so that the policymakers in the future can take into consideration the regulatory impact. So we can't do

that unless we are hearing from you. So please do tell us, and we can take your mandate to actually really engage in conversations like this.

(RD) Thank you very much for that. I think that's a really powerful and useful intervention. Any other comments or thoughts?

So we're almost exactly on time which is extraordinary. I'd like to invite you first to thank our very distinguished panel for their effort and their time and their commitment and their thoughts.

[Applause]

(RD) And I would like to, and I know I am the person who stands between you and a glass of wine or a bottle of beer, I'm hoping there's beer, Sue is giving me the nod there is beer. I'd like to thank a few more people. I'd like that thank Research Councils UK for supporting CREATe; the AHRC, the ESRC and the EPSRC. I'd like to thank Wellcome again for letting us in and for staging this event, their support has really been invaluable and, in particular, I'd like to thank Sue Davies who has now disappeared, perhaps she knew this was coming. Sue has been extraordinary in terms of her support and time and commitment to this project and we wouldn't be here if not for Sue, and that also goes for Victoria Stobo as well, who has been the research assistant on this project and I am sure a lot of you have been liaising with in the run up to this event. I would just like to then thank all of you for coming along, for staying for the day, for engaging in the conversation, and hopefully it is the start of a conversation that we can continue with the support of the TNA and other representative organisations. Thanks very much.

PART III – BIBLIOGRAPHY

BIBLIOGRAPHY

Books and articles

Cook, T., "Archives and the Law" (1984) 18 *Archivaria* 20-24

Deazley, R., and Stobo, V., *Archives and Copyright: Risk and Reform*, CREATe Working Paper 2013/3 (March 2013), available at: http://www.create.ac.uk/wp-content/uploads/2013/04/CREATe-Working-Paper-No-3-v1-1.pdf

Dryden, J., *Copyright in the Real World: Making Archival Material Available on the Internet* (Doctoral Thesis, University of Toronto, 2008), available at: https://tspace.library.utoronto.ca/bitstream/1807/11198/1/Dryden_Jean_E_200806_PhD_thesis.pdf

Hirtle, P.B., Hudson, E., and Kenyon A.T., *Copyright and Cultural Institutions* (New York: Cornell University Library, 2009)

Korn, N., *In from the Cold: An assessment of the scope of 'Orphan Works' and its impact on the delivery of services to the public* (April 2009), available at: http://www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx

Kretschmer, M., and Towse, R. (eds), *What Constitutes Evidence for Copyright Policy?, Digital proceedings of ESRC Symposium* (2012), available at: http://www.copyrightevidence.org/create/esrc_evidence_symposium

Stobo, V., *Copyright and Risk: Scoping the Wellcome Digital Library*, CREATe Working Paper 2013/10 (December 2013), available at: http://www.create.ac.uk/wp-content/uploads/2013/12/CREATe-Working-Paper-No.10.pdf

Legislation and official documents

Copyright Act 1976 (US)

Copyright Designs and Patents Act 1988 c.42 (UK)

Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works (the *Orphan Works Directive*)

Enterprise and Regulatory Reform Act 2013 c.24

Gowers, A., *Gowers Review of Intellectual Property* (December 2006), available at: http://www.official-documents.gov.uk/document/other/0118404830/0118404830.asp

Hargreaves, I., *Digital Opportunity: A Review of Intellectual Property and Growth* (May 2011), available at: http://www.ipo.gov.uk/ipreview-finalreport.pdf

HM Government, *Archives for the 21st Century* (November 2009), available at: http://www.nationalarchives.gov.uk/archives-sector/archives-21-century.htm

HM Government, *Modernising Copyright: A modern, robust and flexible framework. Government response to consultation on copyright exceptions and clarifying copyright law* (December 2012), available at: http://www.ipo.gov.uk/response-2011-copyright-final.pdf

HM Government, *Taking Forward the Gowers Review of Intellectual Property: Proposed Changes to Copyright Exceptions* (November 2007), available at: http://www.ipo.gov.uk/consult-copyrightexceptions.pdf

HM Government, *Taking Forward the Gowers Review of Intellectual Property: Second Stage Consultation on Copyright Exceptions* (December 2009), available at: http://www.ipo.gov.uk/consult-gowers2.pdf

Intellectual Property Office (UK), *Factsheet – Orphan Works Licensing Scheme and Extended Collective Licensing* (August 2013), available at: http://www.ipo.gov.uk/orphanworks-licensing.pdf

Intellectual Property Office (UK), *Orphan Works, Impact Assessment No. BIS1063* (June 2012), available at: http://www.ipo.gov.uk/consult-ia-bis1063-20120702.pdf

The National Archives, *Annual Report and Accounts 2011-12* (London: The Stationary Office, 2012)